# Chance and predictability in evolution : The genomic basis of convergent dietary specializations in an adaptive radiation

Vizueta, Joel

2019-09

Vizueta , J , Macias-Hernandez , N , Arnedo , M A , Rozas , J & Sanchez-Gracia , A 2019 , ' Chance and predictability in evolution : The genomic basis of convergent dietary specializations in an adaptive radiation ' , Molecular Ecology , vol. 28 , no. 17 , pp. 4028-4045 . https://doi.org/10.1111/mec.15199

http://hdl.handle.net/10138/323386
https://doi.org/10.1111/mec.15199

acceptedVersion

MOLECULAR ECOLOGY

# Chance and predictability in evolution: the genomic basis of convergent dietary specializations in an adaptive radiation



| | |
|---|---|
| Journal: | *Molecular Ecology* |
| Manuscript ID | MEC-19-0643.R1 |
| Manuscript Type: | Original Article |
| Date Submitted by the Author: | n/a |
| Complete List of Authors: | Vizueta, Joel; Universitat de Barcelona, Departament de Genètica, Microbiologia i Estadística and Institut de Recerca de la Biodiversitat (IRBio)<br>Macías-Hernández, Nuria; University of Helsinki, Finnish Museum of Natural History<br>Arnedo, Miquel; Universitat de Barcelona, Departament de Genètica, Microbiologia i Estadística and Institut de Recerca de la Biodiversitat (IRBio)<br>Rozas, Julio; Universitat de Barcelona, Departament de Genètica, Microbiologia i Estadística and Institut de Recerca de la Biodiversitat (IRBio)<br>Sánchez-Gracia, Alejandro; Universitat de Barcelona, Departament de Genètica, Microbiologia i Estadística and Institut de Recerca de la Biodiversitat (IRBio) |
| Keywords: | Oceanic islands, Spiders, Diet specialization, Positive selection, Heavy metals, Toxins |
| | |

# Chance and predictability in evolution: the genomic basis of convergent dietary specializations in an adaptive radiation

Joel Vizueta[1], Nuria Macías-Hernández[2, 3], Miquel A. Arnedo[4], Julio Rozas[1]* and Alejandro Sánchez-Gracia[1]*

1 Departament de Genètica, Microbiologia i Estadística, and Institut de Recerca de la Biodiversitat (IRBio), Facultat de Biologia, Universitat de Barcelona, Diagonal 643, 08028, Barcelona, Spain.

2 Laboratory for Integrative Biodiversity Research, Finnish Museum of Natural History, University of Helsinki; PO Box 17, 00014 Helsinki, Finland.

3 Island Ecology and Evolution Research Group, Instituto de Productos Naturales y Agrobiología (IPNA-CSIC). C/Astrofísico Francisco Sánchez 3. La Laguna, Tenerife, Canary Islands, 38206, Spain

4 Departament de Biologia Evolutiva, Ecologia i Ciències Ambientals and Institut de Recerca de la Biodiversitat (IRBio), Facultat de Biologia, Universitat de Barcelona, Diagonal 643, 08028, Barcelona, Spain

* Corresponding authors. E-mail: jrozas@ub.edu, elsanchez@ub.edu

## Abstract

The coexistence of multiple eco-phenotypes in independently assembled communities makes island adaptive radiations the ideal framework to test convergence and parallelism in evolution. In the radiation of the spider genus *Dysdera* in the Canary Islands, species diversification occurs concomitant with repeated events of trophic specialization. These dietary shifts, to feed primarily on woodlice, are accompanied by modifications in morphology (mostly in the mouthparts), behaviour and nutritional physiology. To gain insight into the molecular basis of this adaptive radiation, we performed a comprehensive comparative transcriptome analysis of five Canary Island *Dysdera* endemics representing two evolutionary and geographically independent events of dietary specialization. After controlling for the potential confounding effects of hemiplasy, our differential gene expression and selective constraint analyses identified a number of genetic changes that could be associated with the repeated adaptations to specialized diet of woodlice, including some related to heavy metal detoxification and homeostasis, the metabolism of some important nutrients and venom toxins. Our results shed light on the genomic basis of an extraordinary case of dietary shift convergence associated with species diversification. We uncovered putative molecular substrates of convergent evolutionary changes at different hierarchical levels, including specific genes, genes with equivalent functions, and even particular amino acid positions. This study improves our knowledge of rapid adaptive radiations and provides new insights into the predictability of evolution.


**Keywords:** Oceanic islands, Spiders, Diet specialization, Comparative transcriptomics, Differential gene expression, Positive selection, Heavy metals, Toxins

## Introduction

The current limited knowledge of the evolutionary mechanisms underlying diversification compromises our ability to manage and conserve biodiversity (Mergeay & Santamaria, 2012). Evolutionary biology provides a unifying conceptual framework to successfully identify key diversification drivers through the study of molecular variation. As many other fields, evolutionary biology has fully entered the genomics era, which opens up the possibility of tackling longstanding questions regarding biodiversity in a more fruitful way and at a lower cost (Losos et al., 2013). Although often seen as a gradual process that requires the action of different evolutionary forces acting steadily over long periods of time (Coyne & Orr, 2004), speciation can be very rapid under unstable environmental and ecological conditions. In fact, one of the most promising approaches to disclose the relative impact of these driving forces is the study of species radiations in nature, i.e., the rapid appearance of a high number of species from a single common ancestor (Schluter, 2000). In adaptive radiations, such as the classic examples of Darwin's finches (Almén et al., 2016) and the cichlids in the great lakes of Eastern Africa (Henning & Meyer, 2014), significant morphological differences appear over short periods of time despite the low levels of genetic divergence accumulated at the genomic level. Nevertheless, the relative role of natural selection and of other non-adaptive forces in such relevant evolutionary processes is a matter of scientific debate (Muschick, Indermaur, & Salzburger, 2012).

Oceanic islands are considered natural laboratories for studying evolution. The entire biota of these islands is derived from a few initial colonization events followed by local diversification, which generates high levels of endemism and ecomorphological differentiation (MacArthur & Wilson, 1967; Mayr, 1942; Whittaker & Fernández-Palacios, 2007). Thus, the biota of oceanic islands can be interpreted as the result of successful

independent evolutionary experiments starting with a single or multiple colonization events from the continent (Emerson, 2002). The comparative analysis of such independent events and the subsequent island radiation (both within and between islands) in different archipelagos provides new insights into the general evolutionary process generating biological diversity (Gillespie & Roderick, 2002; Losos & Ricklefs, 2009). Such approximation has been successfully applied in a number of studies on oceanic islands (Losos, Jackman, Larson, Queiroz, & Rodriguez-Schettino, 1998; Stroud & Losos, 2016), such as Hawaii (Gillespie, 2004), the Galapagos (Grant & Grant, 2008) and the Canary Islands and Madeira archipelagos (Juan, Emerson, Oromí, & Hewitt, 2000; Machado, Rodríguez-Expósito, López, & Hernández, 2017), where explicit hypotheses on the evolutionary processes underlying radiations have been tested.

The radiation of the genus *Dysdera* Latreille, 1804 (Araneae: Dysderidae) in the Canary Islands is one of the most spectacular examples of island species diversification within spiders (Arnedo, 2001; Arnedo, Oromí, Múrria, Macías-Hernández, & Ribera, 2007). As many as 47 endemic species of this species-rich Mediterranean genus (approximately 250 species) have been reported in the Canary Islands (Macías-Hernández, López, Roca-Cusachs, Oromí, & Arnedo, 2016; World Spider Catalog, 2019). The spiders of the genus *Dysdera* are active nocturnal hunters that spend the daytime in silk retreats and are usually found under stones, dead logs or leaf litter or even living in caves (Arnedo et al., 2007). This genus stands out among spiders in having evolved trophic specialization; i.e., several species have been shown to feed preferably (facultatively or even obligatorily) on terrestrial woodlice (Crustacea: Isopoda) (Řezáč & Pekár, 2007; Řezáč, Pekár, & Lubin, 2008), a prey rejected by most generalist predators (Pekár, Líznarová, & Řezáč, 2016). Available evidence suggests that prey specialization (i.e., stenophagy) has appeared several times, both on the continent

and on the islands. Interestingly, the morphology of mouth parts predicts both dietary preferences and capture strategy (chelicerae used as pincers, forks or keys) and the frequency of captures among the specialists (Řezáč et al., 2008). All cheliceral types observed in continental species have also evolved repeatedly in the Canary Islands, suggesting that prey segregation is a major driving force of the spectacular diversification of the genus on the islands (Arnedo et al., 2007). Woodlice are a difficult prey for other arthropods because of their morphological, chemical and behavioural defences (Gorvett, 1956; Sutton, 1980). These defences comprise dorsally protective armour, gland secretions producing repulsive odours, indigestibility to many predators, and behavioural patterns such as nocturnal activity, rolling into a ball or adhering to surfaces when threatened (Schmalfuss, 1984; Sutton, 1980). In addition, these organisms accumulate high concentrations of heavy metals from the soil, making them even more toxic to predators (Drobne, 1997). Consequently, woodlice are rarely eaten by generalist predators. Within arthropods, only spiders and ants have developed specialized strategies to feed on this prey (Dejean, 1997; Pekár et al., 2016). Nevertheless, despite all this morphological and experimental evidence, the genetic basis of this remarkable adaptation is completely unknown.

Moreover, the study of the molecular basis of such an extraordinary phenotypic convergence offers an opportunity to address the question of predictability and repeatability of the evolutionary process. Given that it is not possible to rerun the tape of evolution, the study of parallel evolutionary outcomes in different scenarios provides a fairly good framework to ascertain both to what extent similar molecular solutions has been exploited repeatedly, and which aspects are predictable at different hierarchical levels (i.e., at the nucleotide, gene, pathway or function level). Among *Dysdera* spiders, the specialized woodlice eaters (i.e., oniscophagous species) possess, in addition to the morphological modifications of chelicera,

important behavioural and nutritional adaptations to feed on isopods (Hopkin & Martin, 1985; Řezáč & Pekár, 2007; Toft & Macías-Hernández, 2017). With the aim of understanding the genetic basis of these specific adaptations and to shed some light on the longstanding debate of how predictable is molecular evolution, we designed a case study that included adult individuals from two pairs of recently diverged endemic specialist-generalist species from the Canary Islands, likely representing two phylogenetically and geographically independent dietary shifts from a generalist ancestor. Our survey included the GV pair: *Dysdera gomerensis* Strand, 1911 (El Hierro) and. *D. verneaui* Simon, 1883 (Tenerife), the TB pair: *D. tilosensis* Wunderlich, 1992 and. *D. bandamae* Schmidt, 1973 (Gran Canaria), and a third generalist endemic species external to both pairs: *D. silvatica* (La Gomera) (Arnedo pers. Comm; Macías-Hernández, Oromí, & Arnedo, 2008; Vizueta et al., 2017), which was used as an outgroup (Figure 1). We compared the transcriptome profiles and the selective constraint patterns between specialists and generalists to identify the genomic regions responsible for the rapid dietary adaptation of *Dysdera* species in the Canary Islands. We studied transcriptomic data from adult individuals, we were able to detect putative adaptive changes associated with food detection and assimilation, including its digestive and metabolic aspects. True homoplasy can arise by evolving the same (or similar) trait from either a non-shared common ancestor (convergent evolution) or a shared ancestor but through evolutionarily independent events (parallel evolution). Here, we will refer to both cases with the general term of "convergence". We aimed to detect those evolutionary changes required to explain a repeated character state in the two specialist lineages, either a gene expression profile or a selective constraint pattern, matching phenotypic convergence. Nevertheless, both incomplete lineage sorting of (ILS; Maddison, 1997) and species hybridization can produce fundamental discordances between gene trees and the species tree, a phenomenon commonly referred to as "hemiplasy" (Avise & Robinson, 2008), giving rise to the illusion of homoplasy

and the erroneous inference of convergence (Mendes, Hahn, & Hahn, 2016; Wu, Kostyun, Hahn, & Moyle, 2018).

Here, and after controlling for the potential confounding effects of hemiplasy, we identified clear signals of homoplasy at different hierarchical levels likely attributable to adaptive convergence in specialist species. Noticeably, we even find signals of this adaptive process at the amino acid level. The repeated changes matching phenotypic convergence found in this study mostly affected genes and gene functions associated with the strategy of detoxifying heavy metals (and perhaps other toxic substances) accumulated by woodlice, to the enhanced assimilation of some nutrients and, to a lesser extent, to venom composition.

157 # Material and Methods

158 ## Study design and sample materials

159 Our study design included two pairs of phylogenetically related *Dysdera* species endemic
160 from the Canary Islands. Each pair of close relatives was composed of a generalist and a
161 specialist (stenophagous) species regarding their diet and shared a generalist ancestor, which
162 implies that at least two specialization events occurred independently during the divergence
163 of these four species, one on each species pair (Figure 1). Both, the phylogenomic analysis
164 performed here and recent multi-locus based phylogenies including other endemic species of
165 this genus (Arnedo et al. unpublished results) indicate that *D. gomerensis* and *D. verneaui* are
166 true sister taxa, while *D. tilosensis* and *D. bandamae* are very closely related, although is
167 difficult to know if they are each other closest relatives. Similarly, the ancestral state
168 reconstruction supports that the ancestor of the complete Canarian radiation was a generalist,
169 while *D. tilosensis* is a derived specialist from a generalist ancestor. For the case of *D.*
170 *gomerensis* this is much more difficult to establish because of the phylogenetic uncertainty,
171 probably due to a very rapid radiation of these species group. In any case, this rapid radiation
172 however makes that most candidate changes in the *D. gomerensis* lineage (see below), would
173 be adaptations to stenophagy, independently of whether the ancestor was a complete
174 generalist, or just a facultative intermediate.
175
176 The two specialists species of our study show modifications in their mouthparts that have
177 been associated with a preference for using isopods as a prey (Řezáč et al., 2008; Macías-
178 Hernández et al, in prep) (see Figure 1). We collected 16 individuals of *Dysdera tilosensis*
179 (10 males and 6 females) and 14 individuals of *D. bandamae* (5 males and 9 females) in Gran
180 Canaria, and 12 males of *D. verneaui* in Tenerife and 15 females of *D. gomerensis* in El
181 Hierro (Table S1). We also included in the analysis a fifth Canary Island endemic *Dysdera*

species, the generalist *D. silvatica*, as an outgroup and to polarize the evolutionary changes in internal branches (Vizueta et al., 2017) (Figure 1).

**Transcriptomic analysis**

For each species, we sequenced the transcripts from the palps (*PALP*), the first pair of legs (*LEG#1*), all other legs (*LEG#234*), and the rest of the body (*REST*), separately in four different RNAseq experiments. We applied this strategy to maximize the detection of low expressed genes, especially chemosensory gene family members in spider appendices (see Vizueta et al., 2017 and Frías-López et al., 2015; Supplementary Methods). Specimens were starved for two weeks at the laboratory and posteriorly fixed in liquid nitrogen and stored at −80 °C until further processing. From the total RNA, we sequenced the transcriptomes in the Illumina HiSeq 4000 platform using pair-end libraries (100-bp reads; Table S1). A detailed description of raw data pre-processing, transcriptome assembly and functional annotation of the transcripts from the four species is available in Supplementary Methods.

**Species-tree, gene-tree discordance, and risk of hemiplasy**

We identified all groups of homologous genes that share at least one member in the ancestor of the five *Dysdera* species (i.e., orthology groups) using OrthoMCL with default parameters (Li, Stoeckert, & Roos, 2003). We further separated single-copy orthologs from multigene families. Since at the moment of starting this work, all published phylogenetic analyses including the studied species were based on few genes (Arnedo, 2001; Arnedo et al., 2007), we performed a more comprehensive phylogenomic analysis using all single copy orthologs across the five Canarian *Dysdera* species plus *D. crocata* Koch, 1839 (the phylogenetically closest continental species of this genus with available transcriptome data; Fernández, Hormiga, & Giribet, 2014) (Figure 2). Only complete or nearly complete transcripts free of

premature stop codons were included in the analysis. The multiple sequence alignments (MSA) of the CDS of each orthology group were generated with the program T-Coffee (Notredame, Higgins, & Heringa, 2000) and further concatenated in a single MSA using in house Perl scripts. We set the GTRGAMMA substitution model in a partitioned scheme to obtain the maximum likelihood (ML) tree in the software RAxML (Stamatakis, 2014). Model parameters were estimated independently for each single-copy ortholog and node support was obtained after 500 bootstrap replicates.

We approximated the divergence times between the five Canarian *Dysdera* species by fitting the data from single copy orthologs to the unrooted tree topology of the ML tree after excluding *D. crocata*. We set the same substitution model and partition scheme than in the previous RAxML analysis. We used the penalized likelihood method of Sanderson (2002), implemented in the program r8s v1.80, to generate the ultrametric tree and to estimate node ages (Sanderson, 2003). We set a calibration point in the node representing the split of the *D. silvatica* lineage from the rest of lineages (3.4-7.8 Mya range; Macías-Hernández, Bidegaray-Batista, Emerson, Oromí, & Arnedo, 2013).

We also inferred a species tree that incorporates gene-tree uncertainty using ASTRAL (Zhang, Rabiee, Sayyari, & Mirarab, 2018). For that, we first estimated the ML tree of each individual MSA (i.e., a gene tree for each single-copy ortholog) with RAxML (setting the GTRGAMMA substitution model and calculating node support with 1000 bootstrap replicates). Moreover, we estimated the Hemiplasy Risk Factor (HRF) along the phylogeny using the PePo package (Guerrero & Hahn, 2018). For the analysis, we used the species tree inferred with ASTRAL (with branch lengths in $2N_e$ generation units), a very approximate estimate of the population scaled mutation rate in *D. silvatica* ($\theta = 0.011$; estimate obtained

from a short read alignment to the first genome draft of this species; unpublished results), a generation time of 1.5 years, and six different effective population sizes, $N_e$ ($10^3$, 5 x $10^3$, $10^4$, 5 x $10^4$, $10^5$ and $10^6$). Finally, all candidate genes exhibiting resolved discordant topologies (i.e., with bootstrap support ≥ 75% in at least one node producing discordance with the species tree) were excluded for the downstream functional prediction analyses and their interpretation. Finally, we used the $D_{FOIL}$ statistic (Pease & Hahn, 2015) to test for introgression between the specialist lineages in presence of ILS, using both *D. silvatica* or *D. crocata* as outgoups.

**Differential expression analyses**

Differential expression (DE) analyses were performed separately in each generalist-specialist pair (GV and TB pairs; see Figure 1; Supplementary Methods). Raw reads of the RNAseq from each species and body part were mapped back to their own reference CDS and to the CDS of the other species in the pair by using BOWTIE2 version 2.2.3 (Langmead & Salzberg, 2012). Read counts and TMM-normalized FPKMs (i.e., trimmed mean of log-expression ratios-normalized fragments per kb of exon per million reads mapped) were estimated for single-copy genes and multigene families using RSEM 1.2.19 software (Li & Dewey, 2011). To test for genes showing DE between specialists and generalist species, we calculated the negative binomial dispersion of read counts across species pairs of a set of housekeeping (HK) genes with EdgeR version 3.18.1 (Robinson, McCarthy, & Smyth, 2010). We used this dispersion to conduct the DE analysis between specialist and generalist species. We merged all body parts (within a species) to homogenize the differences in the number of REST samples between species pairs. To avoid type I and II errors associated to this merging, especially when gene expression is higher in *REST* relative to legs (both *LEG#1* and *LEG#234*) and *PALP*, we used total read counts from all samples normalized for each library

size to perform differential expression analyses. The *P*-values of these analyses (one per gene) were corrected for the false discovery rate (Benjamini & Hochberg, 1995) (FDR). We considered that a gene is differentially expressed between two species when expression levels are significantly different with a FDR < 0.05.

**Selective constraints analyses**

We used the adaptive Branch-Site Random Effects Likelihood (aBSREL) model implemented in the HyPhy package (Pond, Frost, & Muse, 2005; Smith et al., 2015) to test if positive selection has occurred repeatedly in the same gene in specialist lineages. This method is based on the parameter ω (the ratio of nonsynonymous ($d_N$) to synonymous ($d_S$) substitution rates, $\omega = d_N/d_S$) and allows fitting an optimal number of ω classes to codon sequence alignments of single-copy orthologs in each branch of the phylogeny (Figure 2; Supplementary Methods). Positive selection is inferred when a gene shows codons fitting a class with $\omega > 1$ in a particular lineage. We also tested for relaxation or intensification of the strength of natural selection in these single copy orthologues in specialist lineages using the RELAX framework in HyPhy (Wertheim, Murrell, Smith, Kosakovsky Pond, & Scheffler, 2015). Besides, we applied the Mixed Effects Model of Evolution (MEME) implemented in the HyPhy package (Murrell et al., 2012) to identify individual sites evolving under episodic positive selection (in one or more lineages) in the set of candidates from PCOC analysis (see below). Both methods are based on the same principle of aBSREL of fitting different probabilistic models of the ω parameter distribution, and also inferred positive selection when $\omega > 1$. Finally, we applied the aBSREL model to test for episodic positive selection acting on gene families in specialist lineages. In this case, we used the same workflow as for the single copy orthologs but applying the FastTree program (Price, Dehal, & Arkin, 2010) to approximate a ML tree of each family.

**Convergent amino acid evolution**

To detect convergent amino acid evolution in specialist lineages, we aligned the amino acid sequences of the PS candidates using the software PRANK and applied the method PCOC (Rey, Guéguen, Sémon, & Boussau, 2018) (Profile Change with One Change), a recently developed approach to identify convergent shifts in the amino acid substitution rate across a phylogeny, to each individual MSA. Moreover, we used computer simulations to test the performance of PCOC method with our empirical data. We applied the same species tree, average sequence length and model parameters set in the PCOC analysis of the observed data to simulate sequences both with convergent (2% of sites undergoing convergent amino acid substitutions) and without convergent changes (Rey et al., 2018). Using these simulated sequences, we estimated the false discovery rate (FDR; using simulations without convergence) and true positive rate (TPR; using simulations with convergent amino acid substitutions) associated with this analysis.

**GO enrichment**

We used R and GOstats (Falcon & Gentleman, 2007) to carry out the gene ontology (GO) enrichment analysis and REVIGO (Supek, Bošnjak, Škunca, & Šmuc, 2011) to generate a graphical representation of the results. We also used Blast2GO suite (Conesa et al., 2005) to identify KEGG pathways enriched in the list of candidates (Kanehisa & Goto, 2000). Hypergeometric tests were performed with dhyper function of the R package STATS.

## Results

We constructed 16 RNA-seq datasets (four different body parts in four species) to obtain four new complete *Dysdera* transcriptomes (Table S1). As expected, both the number of species-specific transcripts (from 170,846 to 347,878) and the number of functionally annotated genes differed between species (Table 1), but the transcriptome completeness, measured as the number and integrity of CEG genes, was quite similar (Table S2). Only 30% of the transcripts encoded protein-coding genes; the rest corresponded to either non-coding transcripts or assembly artefacts (Table 1). Furthermore, ~35% of the predicted proteins showed no significant sequence similarity or conserved profiles with known arthropod genes (i.e., putative orphan genes of the *Dysdera* lineage). Among the annotated proteins, most were chelicerate specific, and ~66% of the top BLAST hits matched spider sequences (Figure S1).

We identified a total of 13,947 orthologous groups across the five Canarian *Dysdera* species, of which 7,958 were free of premature stop codons, and 4,539 showed complete sequences in all species (Figure 2). The number of single-copy orthologues across the five species was 9,473, a number that increased to 19,497 in the GV pair and 24,212 in the TB pair (Table S3). The maximum likelihood (ML) tree that included *D. crocata* (2,472 genes; 2,926,723 bases) confirmed the expected phylogenetic relationships (Figure 1), i.e., that *D. silvatica* is sister to the two generalist/specialist sister lineages (GV and TB). We estimated that *D. gomerensis* and *D. verneaui* diverged approximately ∿4.1 Mya, whereas the split between *D. tilosensis* and *D. bandamae* occurred ∿3.1 Mya; the age of the common ancestor of these four lineages dates to ∿4.5 Mya (analysis based on 4,539 genes; Figure 1). These estimates are similar to those obtained in Macías-Hernández et al., (2013).

These very recent divergence times, especially the short internal branch lengths, indicated that hemiplasy might represent an important confounding factor in our inferences of convergent evolution. Indeed, although the species tree estimated with ASTRAL had the same fully supported topology (the local posterior support for each branch was 1) than as the ML tree based on the concatenated MSA, the final normalized quartet score of this species tree (0.65) uncover a high gene tree conflict in our data set. The risk of hemiplasy (HRF) estimated along the species tree obtained with ASTRAL, varied according to the effective population sizes and the examined branch (Figure 3), being small for $N_e \leq 10^4$, high in branches A and C for $N_e \geq 10^5$, and extremely high in all branches for $N_e \geq 10^6$. Given the high fraction of discordant gene trees observed in our data (5,275 out of 7,784 gene trees; 3,666 with high bootstrap support ≥ 0.75 in at least one discordant node) together with HRF estimates, the surveyed species (and their ancestors) would have intermediate to high effective population sizes, in a range of $10^4 < N_e \leq 10^6$. Although only a small fraction of these inconsistencies might really affect our inferences of homoplasy (see discussion), we specifically considered this confounding factor in our study. In contrast, we did not detect the characteristic hallmark of gene flow between extant specialist lineages in the $D_{FOIL}$ analysis of transcripts, neither by analyzing all transcripts separately nor concatenating them in different gene groups (i.e., all transcripts, all candidates, only gene expression, or only positive selection candidates; results not shown; see below for the precise definition of each type of candidate).

**Gene expression changes matching phenotypic convergence: individual gene level**

Despite the sex-ratio bias of the studied samples (Table S1), the PCA analysis of the eight *REST* samples of the specialist *D. tilosensis* sequenced separately (four males and four females), showed no evidence of sex-specific expression (Figure S2), which is in agreement

with the absence of morphological dimorphism between sexes reported for the Eastern Canarian clade of this genus (Macías-Hernández et al., 2008). We found 774 (out of 19,497) and 1,044 (out of 24,212) genes showing differential expression between specialists and generalist species in the GV and TB pairs, respectively (Figure S3; Table S4). Remarkably, 147 genes (out of 193) had patterns of gene expression matching phenotypic convergence, i.e., the expression profiles had the same trend in both species' pairs with the two specialists significantly under- or overexpressed (hereafter referred to as Matching Gene Expression "MGE" candidates); however, in three cases the tree showed discordant genealogies supported by the entire transcript sequence. The final number of MGE candidates (144 genes) is much higher than that expected by a neutral model of gene expression evolution, both when considering all differentially expressed genes (hypergeometric test; $P = 1.3x10^{-67}$) and separating genes over- or underexpressed in specialist lineages ($P = 2.3x10^{-14}$ and $P = 4.2x10^{-121}$, respectively; hypergeometric test). The proportion of genes significantly underexpressed in specialists was higher both in the two species pairs considered separately (68% in GV and 61% in TB) and, to a much greater extent, across the 144 shared DE candidate genes (114 genes; 79%) (Figure 4; Table S4). All MGE candidates except two functionally uncharacterized proteins (OG9619 and OG15050 in *PALP*) and one phosphatase (OG1641 in *LEGS*), were predominantly expressed in *REST*, (Figure 4; Figure S3), and none of them show DE between males and females of *D. tilosensis* in this body part (results not shown). All these findings indicate that DE analyses are reflecting real differences between specialist and generalist species, and not sex or body part-specific features. Yet, we cannot completely rule out that some of the uncovered candidates was a false positive, so they should be considered as promising candidates to be further validated.

Within the biological processes significantly overrepresented (Figure 5a) among MGE candidates, we identified genes involved in the homeostasis of metal ions; catabolism of amino acids, sugars and chitin and activities of enzymes such as phosphatase and hydrolase. The separate analysis according to the direction of gene expression change showed that the 114 MGE candidates downregulated in specialists are significantly enriched in assembly and organization of chromatin, cytoskeleton and other cellular structures (such as the organelles), potential regulation of developmental processes through the smoothened pathway, cell morphogenesis and growth processes, and catabolism of sugars and amino acids. In contrast, the 30 MGE candidates upregulated in specialists are significantly enriched in GO terms associated to the metabolism of steroids, lipids and dicarboxylic acid, the activities of phosphatases and hydrolase, the membrane transport of different substances, and responses to various external stimuli including cellular response to oxidative stress. Other interesting but not GO-enriched functions of the MGE candidates include iron ion binding (a predicted cytochrome P450 protein overexpressed in specialist spiders) and zinc ion binding (mostly represented by various putative zinc finger-containing proteins; Table S4). Furthermore, we also found two putative venom toxins among the 144 MGE candidates, one of which encodes a protein similar to the α-latrocrustatoxin (underexpressed in specialists), while the other is an U32-aranetoxin-Av1a overexpressed in specialists (see Figure S4 and Table S4 for a more detailed functional description of the MGE candidates, including significantly enriched molecular functions).

Our analysis also detected 21 genes specifically expressed in specialists (i.e., with no detectable expression in generalists; referred to as Matching Specialist-specific Expression "MSE" candidates) (Figure 2). Fifteen of these MSE candidates encode proteins with no significant sequence similarity with any entry in the searched databases; the other six cases,

404 which were not enriched in any GO term, encode catalytic activities, such as hydrolases and
405 peptidases, or are associated with zinc ion-binding proteins, likely involved in the regulation
406 of gene expression (Table S4).
407
408 The highly fragmented nature of the transcripts encoding members of the chemosensory gene
409 families prevented the credible assignation of many orthogroups and, therefore, a reliable DE
410 analysis comparing specialists and generalists. Besides, for the few orthogroups that could be
411 assigned, we did not find any concordant DE pattern in specialists. The same negative results
412 were obtained for the other orthogroups that showed DE in the chemosensory appendages
413 (*PALP* and *LEG#1* and *LEG#234*) in the study of Vizueta et al., (2017).
414
415 **Gene expression changes matching phenotypic convergence: gene function level**
416 Apart from the 144 MGE candidates, the group of genes with DE only in one species pair,
417 627 in GV pair and 897 in TB pair, respectively, also shared a significant number of enriched
418 GO terms (70 terms; hypergeometric test, $P = 4.7 \times 10^{-11}$ for all DE genes; $P = 2.2 \times 10^{-23}$ and $P$
419 $= 1.3 \times 10^{-2}$ for under- and overexpressed genes, respectively). Remarkably, some of these GO
420 terms are the same as those overrepresented among the MGE candidates. For the genes
421 underexpressed in specialists, these included chromatin assembly, the organization of cellular
422 components, such as the cytoskeleton or organelles, and cell growth. Other additional
423 functions, such as phosphate metabolism regulation and the apoptotic process involved in
424 morphogenesis, are also shared among these genes. For the genes overexpressed in
425 specialists, the enriched functions shared between species pairs include lipid catabolism,
426 oxidation-reduction process and response to antibiotics (Figure S4 and Table S4).
427

Among the orthogroups with DE only in one species pair but with equivalent functions, we found genes involved in detoxification processes and genes encoding various members of the cytochrome P450 family (most of them overexpressed in specialists, seven and nine different copies in the GV and TB pairs, respectively) or proteins with esterase activity (seven and six of these enzymes in the GV and TB pairs, respectively). Additionally, we found 29 putative venom toxin-encoding genes in the GV pair (eight overexpressed in G) and 34 in the TB pair (26 overexpressed in T). Interestingly, although the encoding genes differed between the two specialists, they had very similar predicted functions, such as astacin-like metalloprotease toxin precursors or aranetoxin-Av1a and latrotoxins, among others (Table S4).

**Positive selection matching phenotypic convergence: individual gene level**

We applied the aBSREL model to estimate the distribution of ω values of all single-copy orthologues with complete sequences and without premature stop codons (7,784 genes; Figure 2; Table S3). This genome-wide analysis uncovered opposite trends between GV and TB pairs; while the overall selective constraints appear to have been relaxed in the *D. tilosensis* lineage, they intensified in the *D. gomerensis* branch (Figure S5). Nevertheless, the analysis of individual genes identified nine genes with significant differences in the selective constraint values shared between the two specialists (or the two generalists) (RELAX framework analysis, FDR of 0.2; Table S5; referred as Matching Functional Constraint "MFC" candidates). Six of these candidates showed the relaxation hallmark in specialists, while the other three showed a significant increase in the selective constraint. We found some overrepresented biological functions among MFC candidates, such as carbohydrate metabolism and homeostasis, neuropeptide signaling, tRNA modification and pyridine metabolism (Figure S4). When we considered not enriched GO terms, the genes with increased functional constrains in specialists encode proteins similar to the membrane

glycoprotein LIG-1, a neuropeptide receptor-like protein, and zinc finger proteins while the genes that have relaxed most in specialist's species encode two zinc finger-like proteins and a hexokinase.

We identified 297 genes with significant evidence of positive selection in specialist lineages, 169 in *D. gomerensis*, 150 in *D. tilosensis* and, remarkably, 22 cases in which positive selection was inferred in both dietary specialists (Figure 2; Table S6; referred to as Matching Positive Selections "MPS" candidates). After excluding five coding regions with discordant genealogies supported by the entire transcript sequence, the number of MPS candidates (17) is clearly greater than that expected by chance (across the 297 genes showing positive selection in specialists; hypergeometric test; $P = 1.5x10^{-8}$). These genes are enriched in biological processes such as germ cell migration and cell death, cell junction assembly and organization, regulation of the immune response or iron ion homeostasis (Figure 5; Figure S4). Interestingly, one of these genes with endopeptidase inhibitor activity encodes a protein with sequence similarity to U24-ctenitoxin-Pn1a, a possible venom toxin related to cysteine proteinase inhibitors.

The PCOC method (Rey et al., 2018) identified convergent shifts in amino acid preferences in 14 out of the 17 MPS candidates (FDR = 0.03%; TPR = 99.7%; Figure 6; Table S6; Figure S6). Furthermore, in five cases, the subsequent MEME analysis indicated that some of the amino acid sites involved in these convergent shifts have also evolved by positive selection (8 amino acid sites; Figure 6). The target genes include i) the U24-ctenitoxin-Pn1a candidate toxin (OG6752 orthogroup; 6 amino acid changes); ii) OG7181, a transcript encoding a protein similar to tectonin (10 amino acid changes, 3 of them  under); iii) OG9641, a transcript encoding a protein involved in response to oxidative stress (3 amino acid changes,

one of them also detected with MEME); iv) OG11255, a gene that encodes a product similar to a mannose receptor (5 amino acid changes, 2 of them also detected with MEME); v) OG13286, a protein likely encoding a sodium channel (1 amino acid change, also detected with MEME); and vi) OG16682, a hydrolase involved in nitrogen compound metabolism (4 amino acid changes, one of them detected with MEME). The analysis also inferred some amino acid substitutions responsible of a convergent shift of preferences in specialists but without evidence of positive selection in OG9529, a putative dehydrogenase and oxidoreductase (4 amino acids) (Figure S6).

**Positive selection matching phenotypic convergence: gene function level**

Although the group of genes under positive selection in only one of the two specialists (147 in GV pair and 138 in TB pair, respectively) did not share more significantly enriched GO terms than expected by change (only three shared GO were enriched in both pairs; hypergeometric test; $P = 0.19$), the number of total GO terms shared by these two groups is greater than expected ($P = 5.3 \times 10^{-75}$ based on the hypergeometric distribution). Among shared GO terms, we found processes and functions such as chitin metabolism (including proteolysis activity), lipid metabolism, metal ion binding (zinc in both pairs, copper in *D. gomerensis* and iron in *D. tilosensis*), and hydrolase and oxidoreductase activities (Figure S4). In addition, we also detected the signature of positive selection in six genes encoding putative venom toxins: four in *D. gomerensis* and two in *D. tilosensis* (Table S6).

The gene family analysis also uncovered the hallmark of positive selection in five gene families affecting both specialist lineages (Figure 2; Table S6). One family (the OG3133 orthologous group), which included sequences without any functional annotation, also showed copy number variation in the two specialists (2 and 3 copies in *D. gomerensis* and *D.*

503 *tilosensis*, respectively, compared to one in the generalist species). The other four gene
504 families encoded proteins with possible functions in chitin metabolism and sequences similar
505 to carbohydrate and zinc ion-binding proteins, hydrolases and other enzymes with catalytic
506 activity. Again, we found a gene family encoding putative venom components (in this case,
507 with no characterized target) among positively selected gene families.
508

## Discussion

The evolution of stenophagy, dietary specialization from a generalist ancestor, most likely involves gene regulatory changes, amino acid replacements in proteins, and/or even copy number variation in gene families. Here, we focused our analysis on the first two issues since comparative transcriptomics based on *de novo* assemblies prevents accurate estimation of changes in gene expression and gains and losses in gene family members. Our approach allows detecting genetic changes in the genes expressed in adults (either in the same gene or in equivalent gene functions) matching the phenotypic convergence observed in dietary specialist *Dysdera*. Nevertheless, it is largely known that hemiplasy can also produce such matching patterns, inducing false evidence of convergent evolution (Mendes et al., 2016; Wu et al., 2018). Indeed, the high level of gene tree discordance caused by ancestral polymorphisms could potentially explain some of the repeated changes identified in *D. gomerensis* and *D. tilosensis*. Nonetheless, some lines of evidence support that most of the candidates reported in this study accumulated convergent changes in specialist lineages. First, for realistic effective population sizes (i.e., $10^4 < N_e \leq 10^5$; these spiders are island endemic predators with likely low census sizes), the probability of observing discordant trees matching the phenotypic convergence is very low (Figure 3). The estimates of the HRF values in branch B under realistic effective population sizes ranged from 0.001 to 0.134 (Figure 3b and 3c). Therefore, the probability of occurrence of ILS on this branch, accompanied by a mutation in the branch A or in an older lineage creating a false pattern of homoplasy, is much lower than that of true homoplasy (Guerrero &, Hahn, 2018). Second, among the total set of discordant gene trees with high bootstrap support, only the 1.69% (62 out of 3,666) yielded resolved topologies that match exactly the one expected from convergence in specialists, which agrees with hemiplasy risk predictions for intermediate effective population sizes. Even so, and to be conservative, we excluded from the

downstream functional prediction analysis all candidates with gene trees included in this 1.69%. This approach, however, may not be suitable for detecting convergent changes in gene expression in specialists. Actually, the assumption that the regulatory regions responsible of the concordant changes in gene expression of candidate genes are completely linked to the transcribed sequence (i.e., both share the same gene tree) may not be correct. Estimates of the recombination rate in these genomes are not available and, more importantly, some of these mutations could be far away from the coding region, even acting in *trans*. In these cases, however, we would expect that gene-tree discordance will be randomly distributed across the genome. We found, by contrast, a clear bias in our candidates towards genes and functions biologically relevant for dietary specialists. Bearing all this in mind, the fixation of convergent genetic changes remains as the most likely explanation for most of the discordant patterns matching phenotypic convergence, even for MGE candidates. Consequently, we demonstrated that our study design, with two evolutionary replicates of the same dietary specialization event, was able to identify potential candidate genes and groups of functionally equivalent genes responsible in part to these remarkable ecological shifts.

A priori, we would expect that the biological functions targeted by selection are related to prey capture and food assimilation, both in digestive and metabolic aspects. Since genetic changes underlying morphological modifications of the specialists' mouthparts likely involve changes in gene expression patterns during development, they were undetectable in our comparative analysis of adult transcriptomes. However, other aspects related to the detection, attack, consumption and digestion of a prey with remarkable behavioural and chemical defences definitely played a crucial role in specialization. Several studies have revealed significant differences in the growth and nutrient extraction efficiencies in specialist *Dysdera* fed on woodlouse, which suggests the existence of metabolic adaptations (Řezáč & Pekár,

2007; Toft & Macías-Hernández, 2017; Macías-Hernández et al., *in prep*.). Toxicity is the most relevant nutritional aspect that makes isopods a prey commonly rejected by most generalist spiders (Hopkin & Martin, 1985). Indeed, isopods accumulate toxic substances, including high concentrations of heavy metals from the soil, especially copper but also zinc, lead and cadmium, in vesicles such as lysosomes (Paoletti & Hassall, 1999). The toxic effects as well as some of the underlying genetic response mechanisms of heavy metals on terrestrial invertebrates have been known for a long time (Janssens, Roelofs, & van Straalen, 2009; Merritt & Bewick, 2017; Migula, Wilczek, & Babczyńska, 2013). Remarkably, our results are in full agreement with the few comparative transcriptomics studies conducted on these types of animals under different metal-stress conditions (e.g. Gomes, Scott-Fordsmand, & Amorim, 2014; Roelofs et al., 2009; Zapata, Tanguy, David, Moraga, & Riquelme, 2009), including in spiders (Li et al., 2016). These studies demonstrate that arthropods exposed to heavy metals show important gene expression changes relative to controls; remarkably, some of the reported gene targets also appear among our MGE candidates or correspond with some of the molecular functions enriched in our list. Some examples include ABC transporters, amiloride-sensitive sodium channels, ATPases, MAP kinases, ubiquitin ligases, histones, members of the cytochrome P450 family and ribosomal proteins (Table S4). These consistent results across different studies on phylogenetically distant species, support the idea of a relatively well-conserved common mechanism for the tolerance of heavy metal toxicity across animals. The old origin of such an evolutionary mechanism validates our approach for identifying the genetic determinants of stenophagy in *Dysdera.*

**Genetic changes matching phenotypic convergence: metal-induced damage or adaptive response to metal stress?**

We found that most MGE candidates were specifically downregulated in specialists and encoded molecular functions involved in cell response, vesicular transport, organization of organelles and cytoskeleton, cilia assembly, or cell adhesion (Table S4). Noticeably, these are the most frequent cell modifications observed in intestinal tissue damage by heavy metals from the diet (e.g., Bednarska et al., 2016; Köhler & Alberti, 1992; Zhang et al., 2001). Indeed, in soil arthropods subjected to heavy-metal stress, midgut cells show evident histological modifications indicative of metal deposition in intracellular granules and gut epithelial degeneration. Although the downregulation pattern observed in specialist *Dysdera* could be the result of a direct stress-induced perturbation of gene expression caused by the high concentration of heavy metals supplied in a woodlouse-rich diet, they might actually be part of an adaptive biological response to excrete metals or other toxic substances more efficiently, thus avoiding their assimilation (Van Straalen & Roelofs, 2005). Consistent with this hypothesis, we observed concordant DE patterns in some MAP kinase pathway members, which participate in an important stress-activated/immune response cascade (Chmielowska-Bąk & Deckert, 2012), and in some ubiquitin ligases, which, among other functions, are involved in the inhibition of cell growth and cycle arrest in response to DNA damage (Cao & Yan, 2012). The adaptive response in specialists would consist of downregulating a set of genes to keep gut epithelial cells in a semi-degenerated functional and structural state that allows enhanced accumulation of heavy metals in granules and very fast and effective intestinal exfoliation and regeneration.

Our analysis also uncovered a number of upregulated MGE and MPS candidates associated with iron, copper and zinc binding and homeostasis, which can also be part of an adaptive mechanism of detoxification in specialist *Dysdera.* Among these candidates, we found amiloride-sensitive sodium channels, membrane ATPases and ABC and dicarboxylate

608 transporters. These proteins are either antiporters for metal cations or are involved in cellular
609 mechanisms for heavy metal vacuolar sequestration (Ahearn, Sterling, Mandal, &
610 Roggenbeck, 2010) or in cellular metal homeostasis and detoxification (e.g., Sooksa-Nguan
611 et al., 2009; Lee et al., 2014). Another set of interesting candidates are the proteins annotated
612 as syntaxin-5-like proteins with a SNARE domain, which are involved in vesicle tethering
613 and fusion associated with copper ion homeostasis (Norgate et al., 2010) and, in addition to
614 being significantly overexpressed in both specialists, also show signals of positive selection
615 in *D. tilosensis*.
616
617 It is well known that heavy metal-associated toxicity is largely due to damage to the oxidative
618 tissue caused by the accumulation of reactive oxygen species in the cell (Schieber & Chandel,
619 2014). Noticeably, among the upregulated MGE candidates (and those regulated in only one
620 of the specialists), we found members of family 3 of the P450 cytochromes, a group of
621 monooxygenases that constitute the largest and most functionally diverse class of insect
622 detoxification enzymes and that have been implicated in the oxidative detoxification of
623 furanocoumarins, alkaloids, plant secondary metabolites and synthetic insecticides (Nelson &
624 Nebert, 2011). Additionally, we identified among the candidates several esterases, a group of
625 proteins with a role in heavy metal and pesticide detoxification that have been used as
626 biomarkers of metal exposure in many organisms, including spiders (Wilczek, Babczyńska,
627 Migula, & Wencelis, 2003). We identified esterases significantly overexpressed in both
628 specialists, although in this case, the orthogroups of *D. gomerensis* and *D. tilosensis* were
629 different, suggesting possible convergence at the functional level rather than at the gene level.
630 Remarkably, two of these esterases also showed a positive selection signal in *D. gomerensis*.
631

We also detected other MGE candidates associated with the metabolism of some essential nutrients, such as proteins with chitin-binding and chitinase activity, and enzymes involved in the metabolism of amino acids, sugars and lipids. Given that most of these candidates were downregulated in specialists, the adaptive advantage could be associated with a reduction in biosynthetic processes to save energy, presumably to dedicate the energy to detoxification processes. However, the presence of some upregulated and positively selected genes among these metabolic candidates indicates that specialists might also have developed an adaptive mechanism to enhance the assimilation and metabolization of some other nutrients present in woodlice but less accessible to other preys.

Finally, it is worth noting that MPS candidates are also significantly enriched in genes related with the immune system. It has been reported that high concentrations of heavy metals negatively affect important processes, such as phagocytosis and chemotaxis, during the generation of the immune response (Boyd, 2010). The footprint of positive selection detected in specialist *Dysdera*, matching phenotypic divergence, might reflect an adaptive mechanism to alleviate the negative immunomodulation effects of heavy metals. In fact, there is evidence that positive selection promoted local adaptation of herbivore insects to heavy metal polluted environments by enhancing immune functions (van Ooik & Rantala, 2010) suggesting the important adaptive character of this system under metal-stress conditions.

**A possible role of venom toxins in the convergent dietary shift**

Stenophagous spiders (e.g., myrmecophagous, termitophagous and araneophagous spiders) show increased venom toxicity to the preferred prey, while related generalists show similar toxicities to all preys (Pekár, Líznarová, Bočánek, & Zdráhal, 2018). The analysis of venom components in stenophagous species indicates that this difference in efficacy is caused by the

presence of prey-specific toxins, suggesting evolutionary adaptations for more effective exploitation of focal prey. Notably, we identified a number of transcripts encoding venom toxins among the MGE candidates, most of which were upregulated in specialists, an opposite pattern to that obtained for the rest of the MGE candidates. Among others, we found candidates encoding astacin-like metalloproteases. Astacins share common features with serralysins, matrix metallo-endopeptidases, and snake venom proteases and might be involved in the proteolytic processing of other venom toxins or even play a role in extra-oral digestion of prey, which could be important in the specialization of Canarian *Dysdera* to woodlice. Interestingly, the MGE candidates encoding astacin-like metalloproteases belonged to different orthogroups in each specialist species, which suggests an additional example of functional convergence through different genes. Our analysis also uncovered other candidates that encode some lesser-known toxins, such as products with sequence similarity to U24-ctenitoxin-Pn1a (presumably a protease inhibitor), pisautoxin-Dm1a (a toxin from the venom of the spider *Dolomedes mizhoanus* with an unknown target), alpha-latrotoxins (which induce massive neurotransmitter release) and aranetoxins (also with an unknown target). Remarkably, we found that among the alpha-latrotoxins, a transcript with similarity to a crustacean-selective component of spider venom (the alpha-latrocrustatoxin; Grishin, 1998), also showed the signature of positive selection, making it a promising candidate for stenophagy. Further research including venom gland-specific transcriptomes and the study of venom toxicity to different preys would be required to shed light on the role of venom in the convergent dietary specialization of *Dysdera*.

**Repeated adaptation to stenophagy in Canarian endemic *Dysdera*: collateral or parallel evolution?**

Here, we uncovered several pieces of evidence supporting the adaptive divergence hypothesis in stenophagous *Dysdera* inhabiting Western Canary Islands. First, the functional annotation of the majority of genes with concordant changes in gene expression between generalist and specialist spiders clearly points towards an active role of these genes in the dietary shift. Second, we detected repeated episodes of positive selection in the same genes (or functionally related group of genes) in the two specialists' lineages. Furthermore, a significant number of MPS candidates showed convergent amino acid preference shifts in the two focal branches, some of which were also inferred to be under positive selection. Altogether, these results provide new significant evidence that species can find the same molecular solutions to adapt predictably to similar ecological niches more often than previously thought (see Marques et al., 2017; Nosil et al., 2018, for other recent examples).

Specialist *Dysdera* may have repeatedly adapted to stenophagy through parallel or collateral evolution. In the first case, convergence would result from the accumulation of the same or similar mutations in evolutionary independent lineages, whereas in the second, selection on either shared ancestral or introgressed variations, would be the responsible of the convergent patterns (Stern, 2013). In recent years, increasing evidence has emerged suggesting the important role of shared genetic variation as a substrate for driving repeated evolution of ecotypes in nature (e.g. Jones et al., 2012; Marques, Meier, & Seehausen, 2019; Schluter & Conte, 2009; Van Belleghem et al., 2018). Our genome-wide HRF and $D_{FOIL}$ analyses point to that most of our candidates originated from parallel independent evolution (i.e., relatively low risk of random ILS and non-significant $D_{FOIL}$ results). On the other hand, in the five positive selection candidates where the individual gene trees were incongruent, the apparent homoplasy could be the result of collateral evolution. Unfortunately, in these cases, current data would not allow to disentangle collateral evolution from random ILS at the individual

gene level. Accordingly, and to avoid reporting candidates with false patterns of homoplasy, we excluded these five genes with discordant topologies, restricting the analysis on the parallel fixation of *de novo* mutations. Further research including polymorphism from whole genome data would be needed to unequivocally establish the relative role of collateral evolution in the convergence observed in these island endemic spiders.

Altogether, our findings suggest that the ecological opportunity provided by the colonization of the Canary Islands facilitated the exploration of multiple adaptive landscapes by *Dysdera* and its diversification on similar peaks (Mahler, Ingram, Revell, & Losos, 2013), providing an exceptional example of repeatability in evolution and shedding light on the genetic determinants of phenotypic convergence (Stroud & Losos, 2016). Besides, our results support the idea that convergence can involve repeated changes at different hierarchical levels (Rosenblum, Parent, & Brandt, 2014). We found convergent changes at the amino acid, gene and gene function levels that would be mostly associated to the excretion and detoxification of heavy metals accumulated in the preferred prey, and some venom components likely related with prey capture. We also demonstrated that natural selection promoted the fixation of some of these changes, confirming the view that adaptive forces are a primary determinant of phenotypic convergence (Storz, 2016). Moreover, our report uncovering repeated genetic changes in pairs of phylogenetically-close taxa, supports the ongoing debate that the probability of shared molecular changes for convergent phenotypes correlates with node age (Conte, Arnegard, Peichel, & Schluter, 2012). Hence, this study not only provide new evidence on the genomic basis of an extraordinary example of a convergent ecological shift in a non-model organism but also offer new insights into the longstanding debate about predictability in evolution.

**Acknowledgements**

We thank to five anonymous reviewers for their useful comments on the manuscript. We also thank Cristina Frías-López for helping with the RNA extractions, and Matthew Hahn for his suggestions and recommendations. This work was supported by the Ministerio de Economía y Competitividad of Spain (CGL2012-36863, CGL2013-45211, CGL2016-75255 and CGL2016-80651) and the Comissió Interdepartamental de Recerca I Innovació Tecnològica of Catalonia, Spain (2014SGR1055 and 2014SGR1604). J.V. was supported by a FPI grant (Ministerio de Economía y Competitividad of Spain, BES-2014-068437). We acknowledge the Cabildos of Tenerife, Gran Canaria and La Gomera, as well as the Garajonay National Park that have granted us collection permits, and often also helped with lodging and logistics during campaigns.

## References


Ahearn, G. A., Sterling, K. M., Mandal, P. K., & Roggenbeck, B. (2010). *Heavy Metal Transport and Detoxification by Crustacean Epithelial Lysosomes*. *Epithelial Transport Physiology*. Totowa, NJ: Humana Press.

Almén, M. S., Lamichhaney, S., Berglund, J., Grant, B. R., Grant, P. R., Webster, M. T., & Andersson, L. (2016). Adaptive radiation of Darwin's finches revisited using whole genome sequencing. *BioEssays*, *38*(1), 14–20. doi:10.1002/bies.201500079

Arnedo, M. (2001). Radiation of the Spider Genus *Dysdera* (Araneae, Dysderidae) in the Canary Islands: Cladistic Assessment Based on Multiple Data Sets. *Cladistics*, *17*(4), 313–353. doi:10.1006/clad.2001.0168

Arnedo, M. A., Oromí, P., Múrria, C., Macías-Hernández, N., & Ribera, C. (2007). The dark side of an island radiation: Systematics and evolution of troglobitic spiders of the genus *Dysdera* Latreille (Araneae:Dysderidae) in the Canary Islands. *Invertebrate Systematics*, *21*(6), 623–660. doi:10.1071/IS07015

Avise, J. C., & Robinson, T. J. (2008). Hemiplasy: A New Term in the Lexicon of Phylogenetics. *Systematic Biology*, *57*(3), 503–507. doi:10.1080/10635150802164587

Bednarska, A. J., Laskowski, R., Pyza, E., Semik, D., Świątek, Z., & Woźnicka, O. (2016). Metal toxicokinetics and metal-driven damage to the gut of the ground beetle *Pterostichus oblongopunctatus*. *Environmental Science and Pollution Research*, *23*(21), 22047–22058. doi:10.1007/s11356-016-7412-8

Benjamini, Y. H., & Hochberg, Y. (1995). Controlling The False Discovery Rate - A Practical And Powerful Approach To Multiple Testing. *Journal of the Royal Statistical Society*, *57*, 289–300. doi:10.2307/2346101

Boyd, R. S. (2010). Heavy Metal Pollutants and Chemical Ecology: Exploring New Frontiers. *Journal of Chemical Ecology*, *36*(1), 46–58. doi:10.1007/s10886-009-9730-5

769 Cao, J., & Yan, Q. (2012). Histone Ubiquitination and Deubiquitination in Transcription,
770 DNA Damage Response, and Cancer. *Frontiers in Oncology*, *2*, 26.
771 doi:10.3389/fonc.2012.00026
772 Chmielowska-Bąk, J., & Deckert, J. (2012). A common response to common danger?
773 Comparison of animal and plant signaling pathways involved in cadmium sensing.
774 *Journal of Cell Communication and Signaling*, *6*(4), 191–204. doi:10.1007/s12079-012-
775 0173-3
776 Conesa, A., Götz, S., García-Gómez, J. M., Terol, J., Talón, M., & Robles, M. (2005).
777 Blast2GO: a universal tool for annotation, visualization and analysis in functional
778 genomics research. *Bioinformatics*, *21*(18), 3674–3676.
779 doi:10.1093/bioinformatics/bti610
780 Conte, G. L., Arnegard, M. E., Peichel, C. L., & Schluter, D. (2012). The probability of
781 genetic parallelism and convergence in natural populations. *Proceedings of the Royal*
782 *Society B: Biological Sciences*, *279*(1749), 5039–5047. doi:10.1098/rspb.2012.2146
783 Coyne, J. A., & Orr, H. A. (2004). *Speciation*. Sunderland: Sinauer Associates.
784 Dejean, A. (1997). Distribution of colonies and prey specialization in the ponerine ant genus
785 *Leptogenys* (Hymenoptera: Formicidae). *Sociobiology*, *29*, 293–299.
786 Drobne, D. (1997). Terrestrial isopods-a good choice for toxicity testing of pollutants in the
787 terrestrial environment. *Environmental Toxicology and Chemistry*, *16*(6), 1159–1164.
788 doi:10.1002/etc.5620160610
789 Emerson, B. C. (2002). Evolution on oceanic islands: molecular phylogenetic approaches to
790 understanding pattern and process. *Molecular Ecology*, *11*(6), 951–966.
791 Falcon, S., & Gentleman, R. (2007). Using GOstats to test gene lists for GO term association.
792 *Bioinformatics*, *23*(2), 257–258. doi:10.1093/bioinformatics/btl567
793 Fernández, R., Hormiga, G., & Giribet, G. (2014). Phylogenomic Analysis of Spiders

Reveals Nonmonophyly of Orb Weavers. *Current Biology*, *24*(15), 1772–1777. doi:10.1016/j.cub.2014.06.035

Frías-López, C., Almeida, F. C., Guirao-Rico, S., Vizueta, J., Sánchez-Gracia, A., Arnedo, M. A., & Rozas, J. (2015). Comparative analysis of tissue-specific transcriptomes in the funnel-web spider *Macrothele calpeiana* (Araneae, Hexathelidae). *PeerJ*, *3*, e1064. doi:10.7717/peerj.1064

Gillespie, R. (2004). Community Assembly Through Adaptive Radiation in Hawaiian Spiders. *Science*, *303*(5656), 356–359. doi:10.1126/science.1091875

Gillespie, R. G., & Roderick, G. K. (2002). Arthropods on Islands: Colonization, Speciation, and Conservation. *Annual Review of Entomology*, *47*(1), 595–632. doi:10.1146/annurev.ento.47.091201.145244

Gomes, S. I. L., Scott-Fordsmand, J. J., & Amorim, M. J. B. (2014). Profiling transcriptomic response of *Enchytraeus albidus* to Cu and Ni: Comparison with Cd and Zn. *Environmental Pollution*, *186*, 75–82. doi:10.1016/j.envpol.2013.11.031

Gorvett, H. (1956). Tegumental glands and terrestrial life in woodlice. *Proceedings of the Zoological Society of London*, *126*(2), 291–314. doi:10.1111/j.1096-3642.1956.tb00439.x

Grant, P. R., & Grant, B. R. (2008). *How and why species multiply : the radiation of Darwin's finches*. Princeton: Princeton University Press.

Grishin, E. V. (1998). Black widow spider toxins: the present and the future. *Toxicon : Official Journal of the International Society on Toxinology*, *36*(11), 1693–701.

Guerrero, R. F., & Hahn, M. W. (2018). Quantifying the risk of hemiplasy in phylogenetic inference. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(50), 12787–12792. doi:10.1073/pnas.1811268115

Henning, F., & Meyer, A. (2014). The Evolutionary Genomics of Cichlid Fishes: Explosive

819 Speciation and Adaptation in the Postgenomic Era. *Annual Review of Genomics and*
820 *Human Genetics*, *15*(1), 417–441. doi:10.1146/annurev-genom-090413-025412
821 Hopkin, S. P., & Martin, M. H. (1985). Assimilation of zinc, cadmium, lead, copper, and iron
822 by the spider *Dysdera crocata*, a predator of woodlice. *Bulletin of Environmental*
823 *Contamination and Toxicology*, *34*(1), 183–187. doi:10.1007/BF01609722
824 Janssens, T. K. S., Roelofs, D., & van Straalen, N. M. (2009). Molecular mechanisms of
825 heavy metal tolerance and evolution in invertebrates. *Insect Science*, *16*(1), 3–18.
826 doi:10.1111/j.1744-7917.2009.00249.x
827 Jones, F. C., Grabherr, M. G., Chan, Y. F., Russell, P., Mauceli, E., Johnson, J., … Kingsley,
828 D. M. (2012). The genomic basis of adaptive evolution in threespine sticklebacks.
829 *Nature*, *484*(7392), 55–61. doi:10.1038/nature10944
830 Juan, C., Emerson, B. C., Oromí, P., & Hewitt, G. M. (2000). Colonization and
831 diversification: towards a phylogeographic synthesis for the Canary Islands. *Trends in*
832 *Ecology & Evolution*, *15*(3), 104–109. doi:10.1016/S0169-5347(99)01776-0
833 Kanehisa, M., & Goto, S. (2000). KEGG: kyoto encyclopedia of genes and genomes. *Nucleic*
834 *Acids Research*, *28*(1), 27–30.
835 Köhler, H.-R., & Alberti, G. (1992). The Effect of Heavy Metal Stress on the Intestine of
836 Diplopods. *Berichte Naturwissenschaftlich-Medizinischer Verein Innsbruck*, *10*, 257–
837 267.
838 Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature*
839 *Methods*, *9*(4), 357–359. doi:10.1038/nmeth.1923
840 Lee, J. Y., Yang, J. G., Zhitnitsky, D., Lewinson, O., & Rees, D. C. (2014). Structural basis
841 for heavy metal detoxification by an Atm1-type ABC exporter. *Science*, *343*(6175),
842 1133–6. doi:10.1126/science.1246489
843 Li, B., & Dewey, C. N. (2011). RSEM: accurate transcript quantification from RNA-Seq data

with or without a reference genome. *BMC Bioinformatics*, *12*, 323. doi:10.1186/1471-2105-12-323

Li, C.-C., Wang, Y., Li, G.-Y., Yun, Y.-L., Dai, Y.-J., Chen, J., & Peng, Y. (2016). Transcriptome Profiling Analysis of Wolf Spider *Pardosa pseudoannulata* (Araneae: Lycosidae) after Cadmium Exposure. *International Journal of Molecular Sciences*, *17*(12), 2033. doi:10.3390/ijms17122033

Li, L., Stoeckert, C. J., & Roos, D. S. (2003). OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Research*, *13*(9), 2178–89. doi:10.1101/gr.1224503

Losos, J. B., Arnold, S. J., Bejerano, G., Brodie, E. D., Hibbett, D., Hoekstra, H. E., … Turner, T. L. (2013). Evolutionary Biology for the 21st Century. *PLoS Biology*, *11*(1), e1001466. doi:10.1371/journal.pbio.1001466

Losos, J. B., & Ricklefs, R. E. (2009). Adaptation and diversification on islands. *Nature*, *457*(7231), 830–836. doi:10.1038/nature07893

Losos, Jackman, Larson, Queiroz, & Rodriguez-Schettino. (1998). Contingency and determinism in replicated adaptive radiations of island lizards. *Science*, *279*(5359), 2115–2118.

MacArthur, R. H., & Wilson, E. O. (1967). *The Theory of Island Biogeography*. Princeton: Princeton University Press.

Machado, A., Rodríguez-Expósito, E., López, M., & Hernández, M. (2017). Phylogenetic analysis of the genus *Laparocerus*, with comments on colonisation and diversification in Macaronesia (Coleoptera, Curculionidae, Entiminae). *ZooKeys*, *651*, 1–77. doi:10.3897/zookeys.651.10097

Macías-Hernández, N., Bidegaray-Batista, L., Emerson, B. C., Oromí, P., & Arnedo, M. A. (2013). The imprint of geologic history on within-island diversification of woodlouse-hunter spiders (Araneae, Dysderidae) in the Canary Islands. *The Journal of Heredity*,

*104*(3), 341–356. doi:10.1093/jhered/est008

Macías-Hernández, N., López, S. de la C., Roca-Cusachs, M., Oromí, P., & Arnedo, M. A. (2016). A geographical distribution database of the genus *Dysdera* in the Canary Islands (Araneae, Dysderidae). *ZooKeys*, (625), 11–23. doi:10.3897/zookeys.625.9847

Macías-Hernández, N., Oromí, P., & Arnedo, M. A. (2008). Patterns of diversification on old volcanic islands as revealed by the woodlouse-hunter spider genus *Dysdera* (Araneae, Dysderidae) in the eastern Canary Islands. *Biological Journal of the Linnean Society*, *94*(3), 589–615. doi:10.1111/j.1095-8312.2008.01007.x

Maddison, W. P. (1997). Gene Trees in Species Trees. *Systematic Biology*, *46*(3), 523–536. doi:10.1093/sysbio/46.3.523

Mahler, D. L., Ingram, T., Revell, L. J., & Losos, J. B. (2013). Exceptional Convergence on the Macroevolutionary Landscape in Island Lizard Radiations. *Science*, *341*(6143), 292–295. doi:10.1126/science.1232392

Marques, D. A., Meier, J. I., & Seehausen, O. (2019). A Combinatorial View on Speciation and Adaptive Radiation. *Trends in Ecology & Evolution*. doi:10.1016/j.tree.2019.02.008

Marques, D. A., Taylor, J. S., Jones, F. C., Di Palma, F., Kingsley, D. M., & Reimchen, T. E. (2017). Convergent evolution of SWS2 opsin facilitates adaptive radiation of threespine stickleback into different light environments. *PLOS Biology*, *15*(4), e2001627. doi:10.1371/journal.pbio.2001627

Mayr, E. (1942). *Systematics and the Origins of Species*. New York: Columbia University Press.

Mendes, F. K., Hahn, Y., & Hahn, M. W. (2016). Gene Tree Discordance Can Generate Patterns of Diminishing Convergence over Time. *Molecular Biology and Evolution*, *33*(12), 3299–3307. doi:10.1093/molbev/msw197

Mergeay, J., & Santamaria, L. (2012). Evolution and Biodiversity: the evolutionary basis of

biodiversity and its potential for adaptation to global change. *Evolutionary Applications*, *5*(2), 103–106. doi:10.1111/j.1752-4571.2011.00232.x

Merritt, T. J. S., & Bewick, A. J. (2017). Genetic Diversity in Insect Metal Tolerance. *Frontiers in Genetics*, *8*, 172. doi:10.3389/fgene.2017.00172

Migula, P., Wilczek, G., & Babczyńska, A. (2013). *Effects of Heavy Metal Contamination*. *Spider Ecophysiology*. Berlin, Heidelberg: Springer.

Murrell, B., Wertheim, J. O., Moola, S., Weighill, T., Scheffler, K., & Kosakovsky Pond, S. L. (2012). Detecting Individual Sites Subject to Episodic Diversifying Selection. *PLoS Genetics*, *8*(7), e1002764. doi:10.1371/journal.pgen.1002764

Muschick, M., Indermaur, A., & Salzburger, W. (2012). Convergent Evolution within an Adaptive Radiation of Cichlid Fishes. *Current Biology*, *22*(24), 2362–2368. doi:10.1016/J.CUB.2012.10.048

Nelson, D. R., & Nebert, D. W. (2011). *Cytochrome P450 (CYP) Gene Superfamily*. *Encyclopedia of Life Sciences*. Cichester : John Wiley & Sons.

Norgate, M., Southon, A., Greenough, M., Cater, M., Farlow, A., Batterham, P., … Camakaris, J. (2010). Syntaxin 5 Is Required for Copper Homeostasis in *Drosophila* and Mammals. *PLoS ONE*, *5*(12), e14303. doi:10.1371/journal.pone.0014303

Nosil, P., Villoutreix, R., de Carvalho, C. F., Farkas, T. E., Soria-Carrasco, V., Feder, J. L., … Gompert, Z. (2018). Natural selection and the predictability of evolution in Timema stick insects. *Science*, *359*(6377), 765–770. doi:10.1126/science.aap9125

Notredame, C., Higgins, D. G., & Heringa, J. (2000). T-Coffee: A novel method for fast and accurate multiple sequence alignment. *Journal of Molecular Biology*, *302*(1), 205–17. doi:10.1006/jmbi.2000.4042

Paoletti, M. G., & Hassall, M. (1999). Woodlice (Isopoda: Oniscidea): their potential for assessing sustainability and use as bioindicators. *Agriculture, Ecosystems &*

*Environment*, *74*(1–3), 157–165. doi:10.1016/S0167-8809(99)00035-3

Pease, J. B., & Hahn, M. W. (2015). Detection and Polarization of Introgression in a Five-Taxon Phylogeny. *Systematic Biology*, *64*(4), 651–662. doi:10.1093/sysbio/syv023

Pekár, S., Líznarová, E., & Řezáč, M. (2016). Suitability of woodlice prey for generalist and specialist spider predators: a comparative study. *Ecological Entomology*, *41*(2), 123–130. doi:10.1111/een.12285

Pekár, Stano, Líznarová, E., Bočánek, O., & Zdráhal, Z. (2018). Venom of prey-specialized spiders is more toxic to their preferred prey: A result of prey-specific toxins. *Journal of Animal Ecology*, *87*(6), 1639–1652. doi:10.1111/1365-2656.12900

Pond, S. L. K., Frost, S. D. W., & Muse, S. V. (2005). HyPhy: hypothesis testing using phylogenies. *Bioinformatics*, *21*(5), 676–679. doi:10.1093/bioinformatics/bti079

Price, M. N., Dehal, P. S., & Arkin, A. P. (2010). FastTree 2 – Approximately Maximum-Likelihood Trees for Large Alignments. *PLoS ONE*, *5*(3), e9490. doi:10.1371/journal.pone.0009490

Rey, C., Guéguen, L., Sémon, M., & Boussau, B. (2018). Accurate Detection of Convergent Amino-Acid Evolution with PCOC. *Molecular Biology and Evolution*, *35*(9), 2296–2306. doi:10.1093/molbev/msy114

Řezáč, M., & Pekár, S. (2007). Evidence for woodlice-specialization in *Dysdera* spiders: behavioural versus developmental approaches. *Physiological Entomology*, *32*(4), 367–371. doi:10.1111/j.1365-3032.2007.00588.x

Řezáč, M., Pekár, S., & Lubin, Y. (2008). How oniscophagous spiders overcome woodlouse armour. *Journal of Zoology*, *275*(1), 64–71. doi:10.1111/j.1469-7998.2007.00408.x

Robinson, M. D., McCarthy, D. J., & Smyth, G. K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, *26*(1), 139–140. doi:10.1093/bioinformatics/btp616

Roelofs, D., Janssens, T. K. S., Timmermans, M. J. T. N., Nota, B., MariËn, J., Bochdanovits, Z., … Van Straalen, N. M. (2009). Adaptive differences in gene expression associated with heavy metal tolerance in the soil arthropod *Orchesella cincta*. *Molecular Ecology*, *18*(15), 3227–3239. doi:10.1111/j.1365-294X.2009.04261.x

Rosenblum, E. B., Parent, C. E., & Brandt, E. E. (2014). The Molecular Basis of Phenotypic Convergence. *Annual Review of Ecology, Evolution, and Systematics*, *45*(1), 203–226. doi:10.1146/annurev-ecolsys-120213-091851

Sanderson, M. J. (2003). r8s: inferring absolute rates of molecular evolution and divergence times in the absence of a molecular clock. *Bioinformatics*, *19*(2), 301–302. doi:10.1093/bioinformatics/19.2.301

Sanderson, Michael J. (2002). Estimating absolute rates of molecular evolution and divergence times: a penalized likelihood approach. *Molecular Biology and Evolution*, *19*(1), 101–9.

Schieber, M., & Chandel, N. S. (2014). ROS function in redox signaling and oxidative stress. *Current Biology : CB*, *24*(10), R453-462. doi:10.1016/j.cub.2014.03.034

Schluter, D. (2000). *The ecology of adaptive radiation*. Oxford: Oxford Univ. Press, Oxford.

Schluter, D., & Conte, G. L. (2009). Genetics and ecological speciation. *Proceedings of the National Academy of Sciences of the United States of America*, *106 Suppl 1*(Supplement 1), 9955–62. doi:10.1073/pnas.0901264106

Schmalfuss, H. (1984). Eco-morphological strategies in terrestrial isopods. *Symposium of the Zoological Society of London*, *53*, 49–63.

Smith, M. D., Wertheim, J. O., Weaver, S., Murrell, B., Scheffler, K., & Kosakovsky Pond, S. L. (2015). Less Is More: An Adaptive Branch-Site Random Effects Model for Efficient Detection of Episodic Diversifying Selection. *Molecular Biology and Evolution*, *32*(5), 1342–1353. doi:10.1093/molbev/msv022

969 Sooksa-Nguan, T., Yakubov, B., Kozlovskyy, V. I., Barkume, C. M., Howe, K. J.,
970 Thannhauser, T. W., … Vatamaniuk, O. K. (2009). *Drosophila* ABC transporter,
971 DmHMT-1, confers tolerance to cadmium. DmHMT-1 and its yeast homolog, SpHMT-
972 1, are not essential for vacuolar phytochelatin sequestration. *The Journal of Biological*
973 *Chemistry*, *284*(1), 354–62. doi:10.1074/jbc.M806501200
974 Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis
975 of large phylogenies. *Bioinformatics*, *30*(9), 1312–3. doi:10.1093/bioinformatics/btu033
976 Stern, D. L. (2013). The genetic causes of convergent evolution. *Nature Reviews Genetics*,
977 *14*(11), 751–764. doi:10.1038/nrg3483
978 Storz, J. F. (2016). Causes of molecular convergence and parallelism in protein evolution.
979 *Nature Reviews Genetics*, *17*(4), 239–250. doi:10.1038/nrg.2016.11
980 Stroud, J. T., & Losos, J. B. (2016). Ecological Opportunity and Adaptive Radiation. *Annual*
981 *Review of Ecology, Evolution, and Systematics*, *47*(1), 507–532. doi:10.1146/annurev-
982 ecolsys-121415-032254
983 Supek, F., Bošnjak, M., Škunca, N., & Šmuc, T. (2011). REVIGO summarizes and visualizes
984 long lists of gene ontology terms. *PloS One*, *6*(7), e21800.
985 doi:10.1371/journal.pone.0021800
986 Sutton, S. L. (1980). *Woodlice*. New York: Pergamon Press.
987 Toft, Sø., & Macías-Hernández, N. (2017). Metabolic adaptations for isopod specialization in
988 three species of *Dysdera* spiders from the Canary Islands. *Physiological Entomology*,
989 *42*(2), 191–198. doi:10.1111/phen.12192
990 Van Belleghem, S. M., Vangestel, C., De Wolf, K., De Corte, Z., Möst, M., Rastas, P., …
991 Hendrickx, F. (2018). Evolution at two time frames: Polymorphisms from an ancient
992 singular divergence event fuel contemporary parallel evolution. *PLOS Genetics*, *14*(11),
993 e1007796. doi:10.1371/journal.pgen.1007796

van Ooik, T., & Rantala, M. J. (2010). Local Adaptation of an Insect Herbivore to a Heavy Metal Contaminated Environment. *Annales Zoologici Fennici*, *47*(3), 215–222. doi:10.5735/086.047.0306

Van Straalen, N. M., & Roelofs, D. (2005). Cadmium tolerance in a soil arthropod a model of real-time microevolution. *Entomologische Berichten*, *65*(4), 105–111.

Vizueta, J., Frías-López, C., Macías-Hernández, N., Arnedo, M. A., Sánchez-Gracia, A., & Rozas, J. (2017). Evolution of chemosensory gene families in arthropods: Insight from the first inclusive comparative transcriptome analysis across spider appendages. *Genome Biology and Evolution*, *9*(1), 178–196. doi:10.1093/gbe/evw296

Wertheim, J. O., Murrell, B., Smith, M. D., Kosakovsky Pond, S. L., & Scheffler, K. (2015). RELAX: detecting relaxed selection in a phylogenetic framework. *Molecular Biology and Evolution*, *32*(3), 820–832. doi:10.1093/molbev/msu400

Whittaker, R. J., & Fernández-Palacios, J. M. (2007). *Island biogeography: ecology, evolution, and conservation*. Oxford: Oxford Univ. Press.

Wilczek, G., Babczyńska, A., Migula, P., & Wencelis, B. (2003). Activity of Esterases as Biomarkers of Metal Exposure in Spiders from the Metal Pollution Gradient. *Polish Journal of Environmental Studies*, *12*(6), 765–771.

World Spider Catalog. (2019). World Spider Catalog. Version 20.0. *Natural History Museum Bern*, online at http://wsc.nmbe.ch. doi:10.24436/2

Wu, M., Kostyun, J. L., Hahn, M. W., & Moyle, L. C. (2018). Dissecting the basis of novel trait evolution in a radiation with widespread phylogenetic discordance. *Molecular Ecology*, *27*(16), 3301–3316. doi:10.1111/mec.14780

Zapata, M., Tanguy, A., David, E., Moraga, D., & Riquelme, C. (2009). Transcriptomic response of *Argopecten purpuratus* post-larvae to copper exposure under experimental conditions. *Gene*, *442*(1–2), 37–46. doi:10.1016/J.GENE.2009.04.019

Zhang, C., Rabiee, M., Sayyari, E., & Mirarab, S. (2018). ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees. *BMC Bioinformatics*, *19*(S6), 153. doi:10.1186/s12859-018-2129-y

Zhang, Y., Lambiase, S., Fasola, M., Gandini, C., Grigolo, A., & Laudani, U. (2001). Mortality and tissue damage by heavy metal contamination in the German cockroach, *Blattella germanica* (Blattaria, Blattellidae). *Italian Journal of Zoology*, *68*(2), 137–145. doi:10.1080/11250000109356398

## Authors' contributions

A.S-G. and J.R. designed, conceived and supervised the research; N.M-H and M.A.A. provided the biological material. J.V. performed the experiments and the bioinformatics work, and analysed the data. M.A.A. performed the dissecting analysis and participated in the data interpretation. J.V., J.R. and A.S-G. wrote the first version of the manuscript. N.M-H. and M.A.A. revised the manuscript and participated in the writing of the final version. All authors read and approved the final version of the manuscript.

## Data accessibility

The raw sequence data generated for this work has been deposited at the Sequence Read Archive (SRA) under Bioproject PRJNA437566. Additional data and analysis generated in this study have been deposited in Figshare (https://doi.org/10.6084/m9.figshare.7726508.v1).

## Competing interests

The authors declare that they have no competing interests

# Figures

**Figure 1. a.** Map of the Canary Islands showing the geographic location of capture localities. **b.** Phylogenetic relationships and divergence times (scale bar) among surveyed *Dysdera* species. The continental species *D. crocata* was used to root the tree. **c.** Dissecting scope images of the left chelicera: A-B: *Dysdera silvatica* female, La Gomera, A, ventral view; B, lateral view; C-D: *D. verneaui* female, Tenerife, C, ventral view, D, lateral view; E-F: *D. bandamae* female, Gran Canaria, E, ventral view, F, lateral view; G-H: *D. gomerensis* female, La Gomera, G, ventral view, H, lateral view; I-J: *D. tilosensis* male, Gran Canaria, I, lateral view, J, lateral view. Bars indicate the relative lengths of the different parts of the chelicerae to highlight differences between the standard (generalists) and elongated or slightly elongated (specialists) chelicerae. White bar: total length of the basal segment (b), dotted part: length of the cheliceral groove (g). Black bar: length of the cheliceral fang (f). In standard chelicerae, g is approximately 1/3 of b, and f is similar to the distance between the base of the segment and the end of the internal keel (k), while in elongated chelicerae, g is longer than 2/5 of f, and f is longer than k. Scale bar in mm. **d.** Live images of the target *Dysdera* species; photo credit: P. Oromí.

**Figure 2.** Core analyses workflow applied in this study, including a summary of the most relevant results. DE, differential expression; DFC, differential functional constraints; PS, positive selection; *, patterns matching the observed phenotypic convergence.

**Figure 3.** Species tree inferred with Astral showing the risk of hemiplasy along the phylogeny. Hemiplasy risk factor values (HRF) were estimated for all internal branches of the tree. The relative probabilities of hemiplasy and homoplasy were inferred under different

effective population sizes ($N_e$; panels **a** to **d**) and assuming a fixed mutation rate μ per $2N_e$ generations ($2N_e\mu = 5.5 \times 10^{-3}$). HRF values estimated for all internal branches (in brackets) represent the proportion of discordant traits associated with a branch due to hemiplasy.

**Figure 4.** Heat map with body part-specific gene expression profiles of the 144 MGE candidates.

**Figure 5.** Bar charts with the most relevant results of the GO enrichment analyses (see Figure S3 for more detailed versions). **a.** Orthogroups with differential expression profiles matching phenotypic convergence (144 MGE candidates) **b.** Orthogroups under positive selection in the two specialists (17 MPS candidates) **c.** Most representative candidates encoding venom toxins in stenophagous *Dysdera*. Dark and light tones represent the proportion of genes with a given associated GO in the candidate and the population (whole transcriptome) set, respectively.

**Figure 6.** Relevant orthogroups showing evidence of convergent amino acid substitutions. (**a**) orthogroup encoding the venom toxin OG6752. (**b-f**) orthogroups with positions evolving under positive selection. Amino acid positions are shaded with different tones according to their profiles, and only positions with a *PP* equal to or greater than 0.99 according to the PCOC, PC or OC model are shown (Rey et al., 2018). Stars highlight the sites identified as being positively selected in MEME.

## Tables

**Table 1.** Summary of dietary habits, sampling localities, RNA-seq data and assembly statistics for each surveyed *Dysdera* species.

# Supplementary material

## Supplementary figures

**Figure S1.** Distribution of blastx hits across species. Distribution of the top 5 hits from the blastx searches with the transcripts of each *Dysdera* species against the ArthropodDB database.

**Figure S2**. Principal component analysis (PCA) of gene expression profiles of individual *REST* samples from *D. tilosensis*.

**Figure S3.** Venn diagrams showing (**a**) the number of shared genes between species pairs. Differential expressed (DE) genes are showed in brackets; (**b**) the number of DE genes between species pairs and groups of tissues (*LEGS-PALP* refers to the *LEG#1*, *LEG#234* and *PALP*); (**c**) number of MGE candidates across tissues.

**Figure S4.** Tree maps with detailed GO enrichment results generated with REVIGO.

**Figure S5.** Box plots showing the distribution of ω values for all single-copy orthogroups in specialist (orange) and generalist (blue) species.

**Figure S6.** Orthogroups with evidence of convergent amino acid evolution. Amino acid positions are coloured according to their profiles, and only positions with a *PP* equal to or greater than 0.99 according to the PCOC, PC or OC model are shown. Yellow stars highlight the sites identified as positively selected in MEME.

**Supplementary tables**

**Table S1.** RNA-seq statistics.

**Table S2.** Distribution of the percentage of CEG length covered by blastx hits.

**Table S3.** Orthogroups classification.

**Table S4.** List of genes with concordant differential expression profiles between generalist and specialists species.

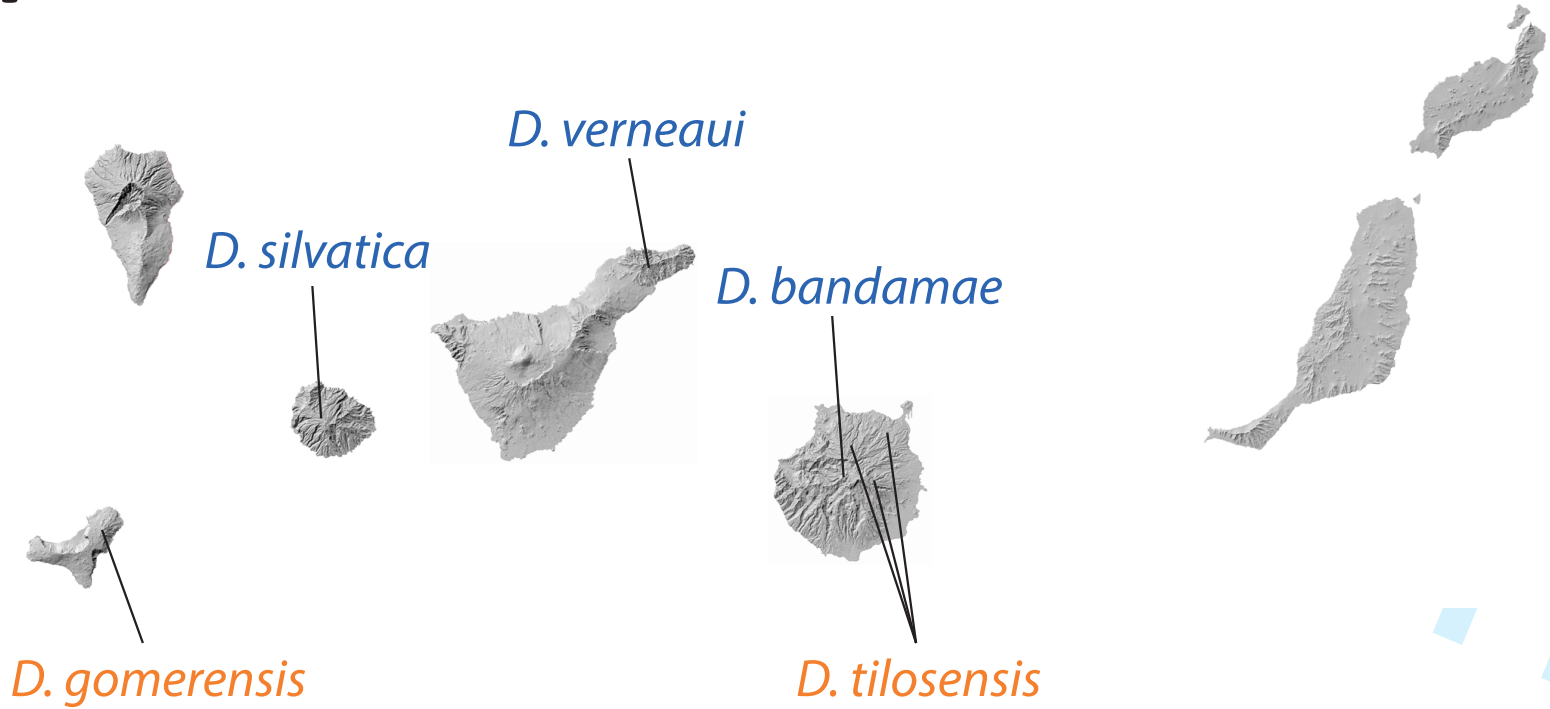
**Table S5.** List of genes with concordant differential functional constraint profiles between generalist and specialist species.

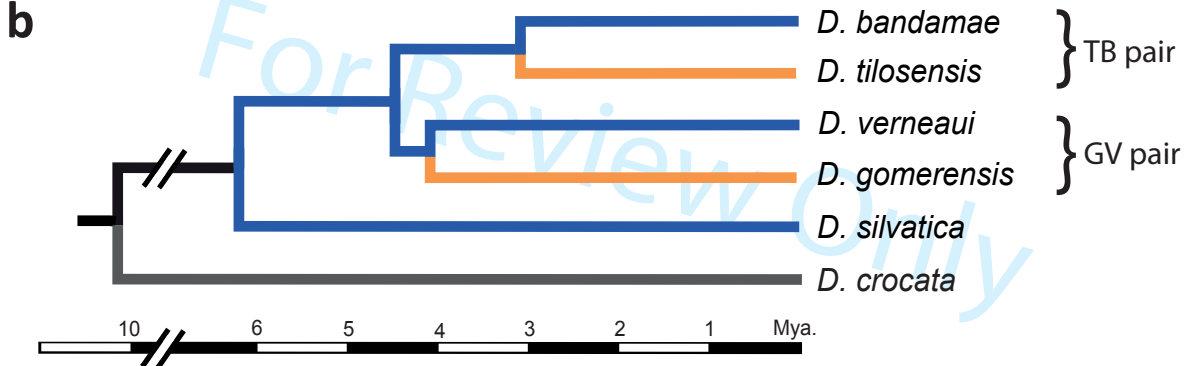
**Table S6.** List of genes with concordant signals of positive selection in specialist species.

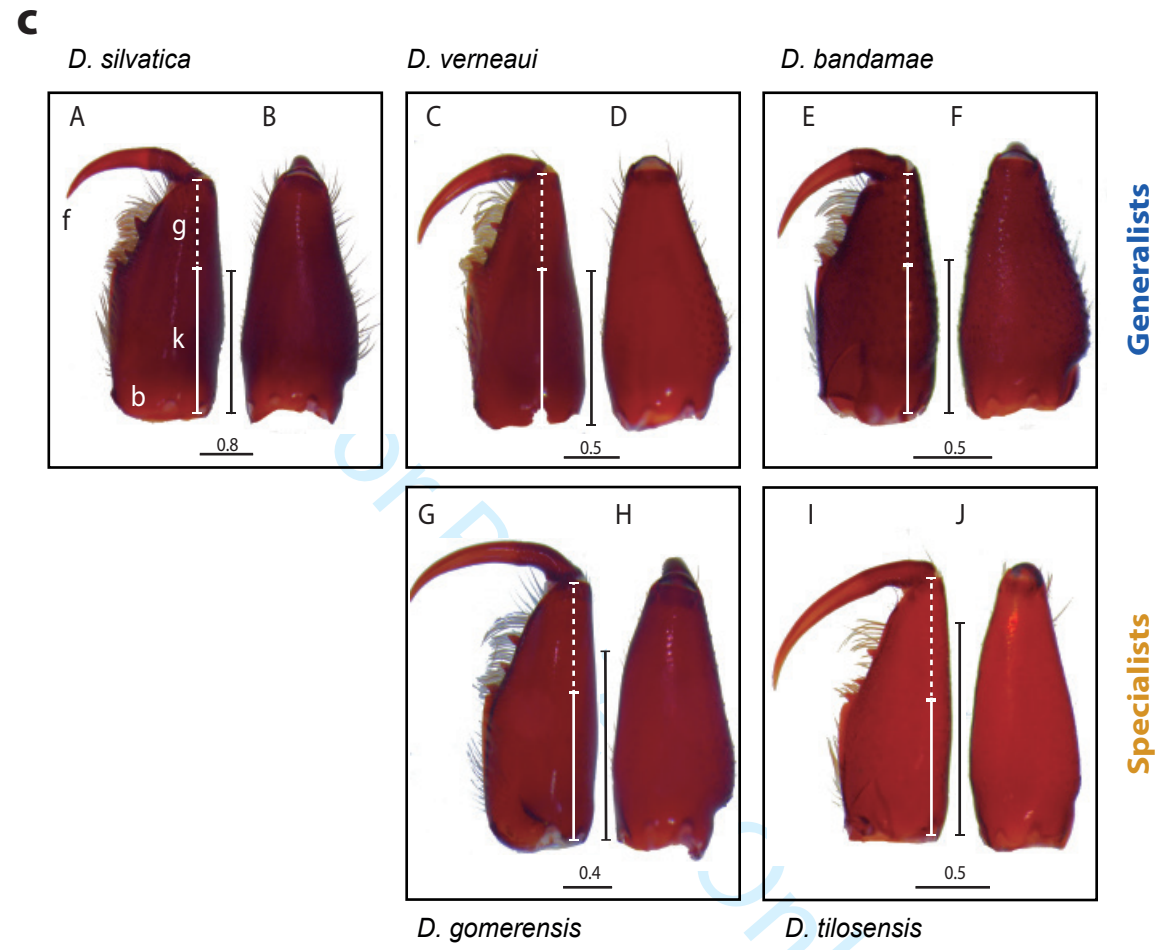
**Supplementary methods**

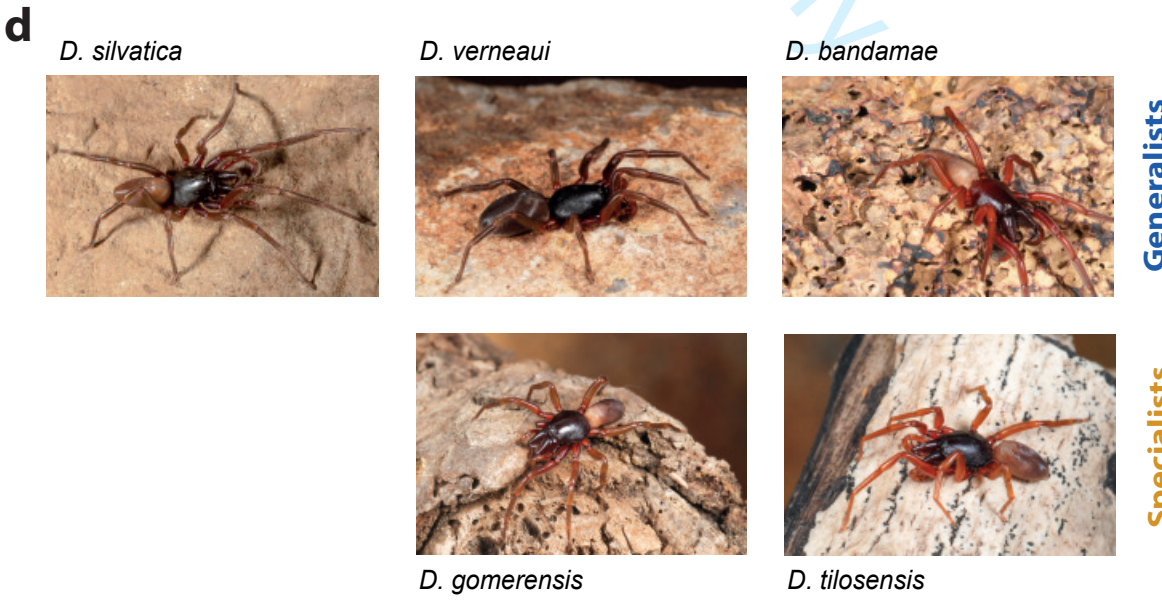
Supplementary Information of transcriptome, differential gene expression and selective constraint analyses.

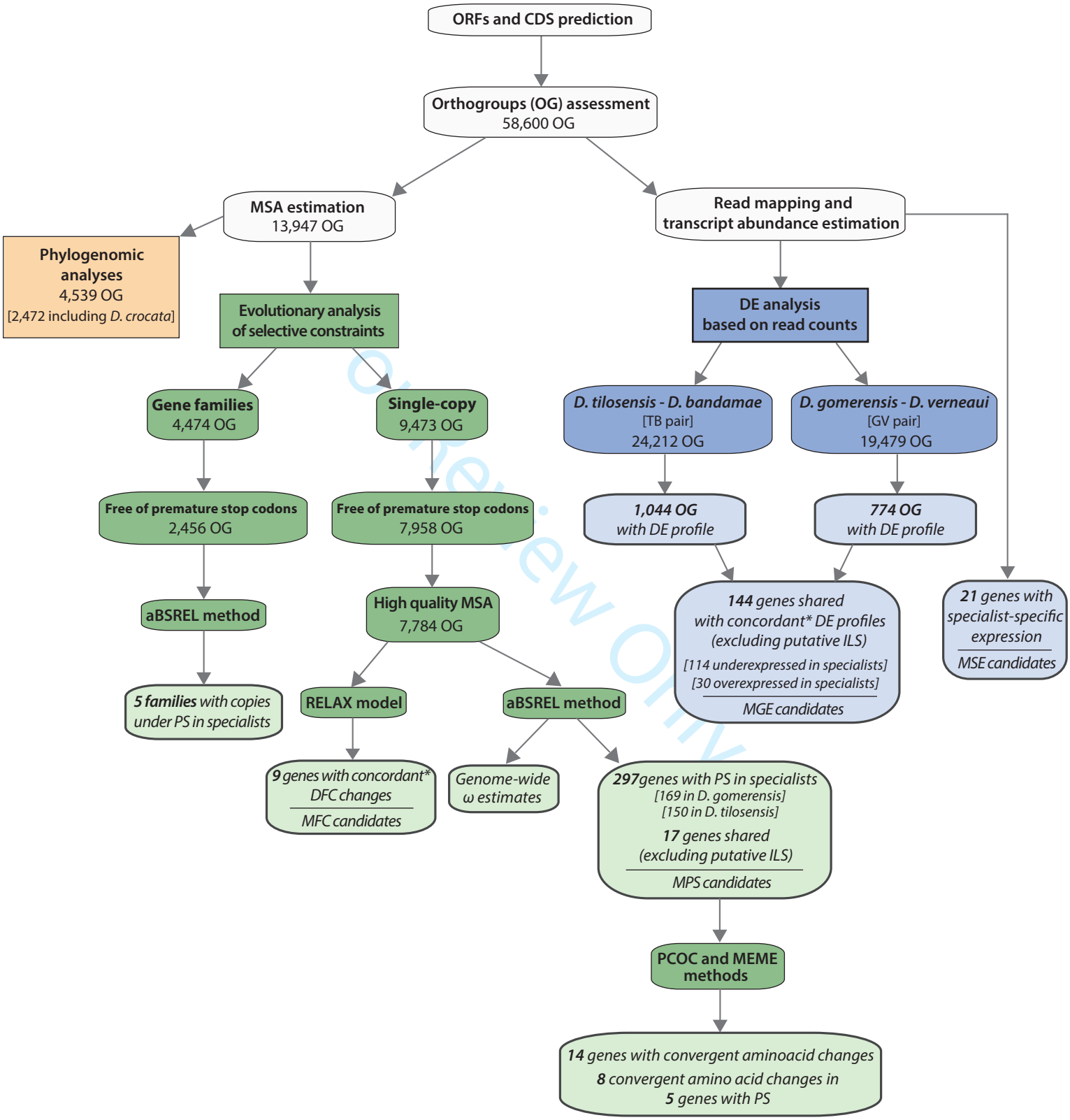
**a**

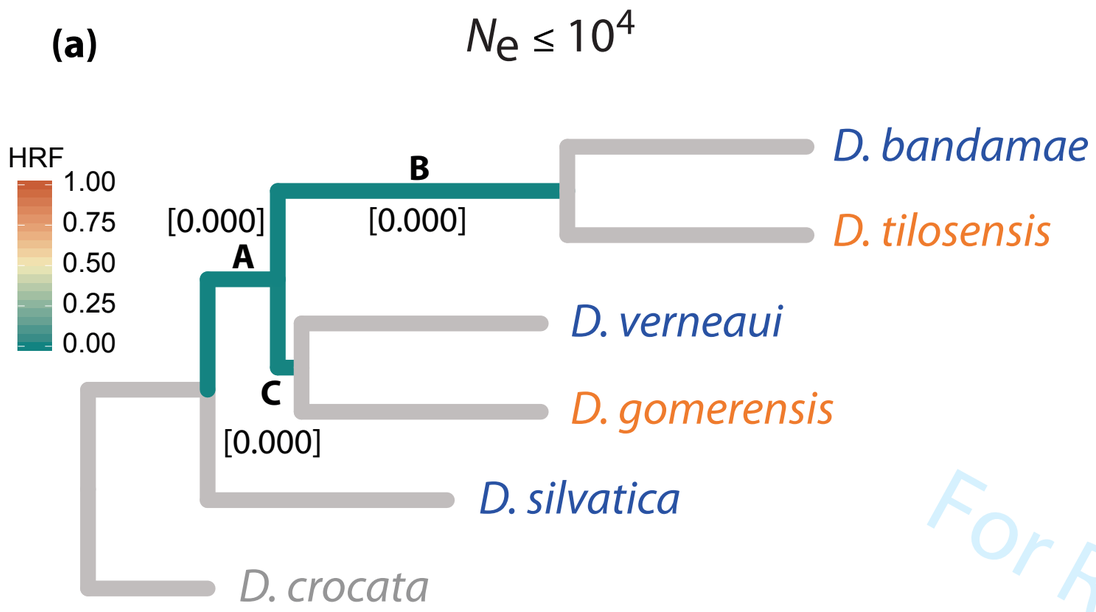

D. verneaui
D. silvatica
D. bandamae
D. gomerensis
D. tilosensis

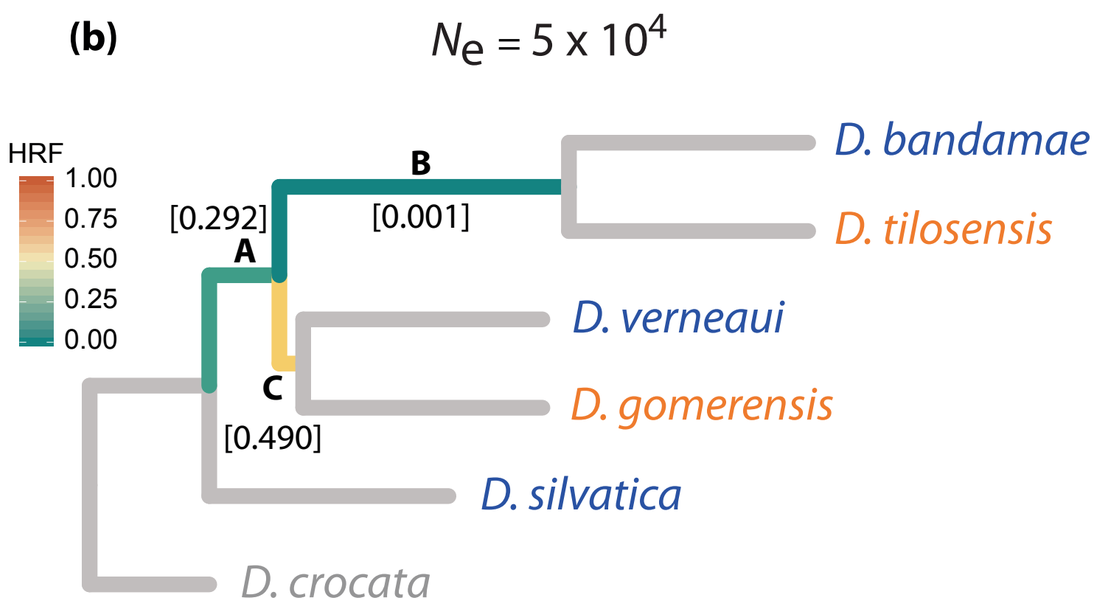


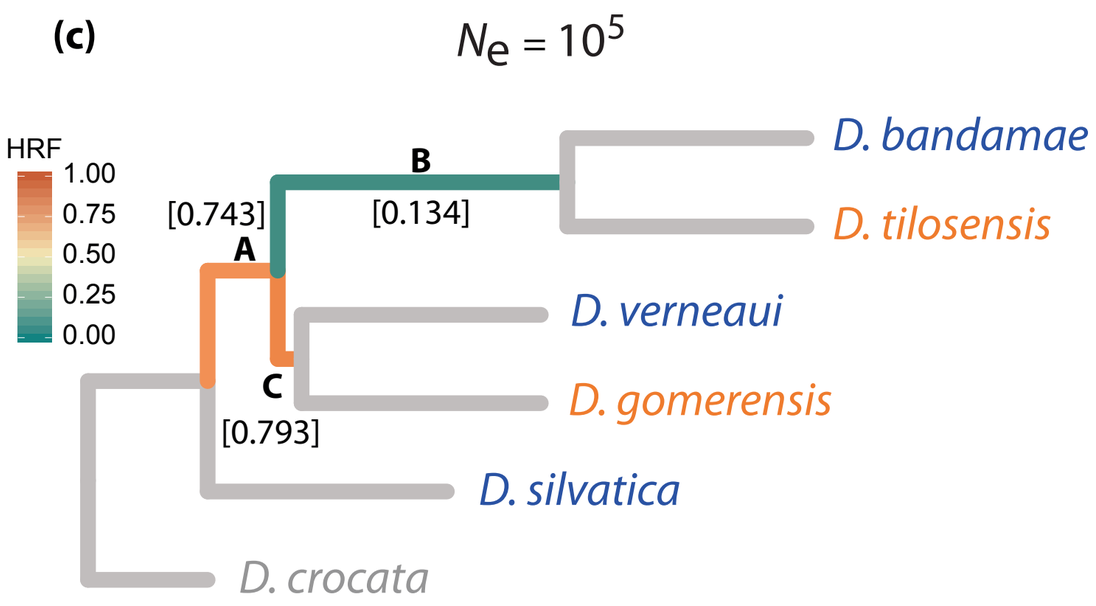

c
D. silvatica
D. verneaui
D. bandamae
A
B
C
D
E
F
f
g
k
b
0.8
0.5
0.5
Generalists
G
H
I
J
0.4
0.5
Specialists
D. gomerensis
D. tilosensis

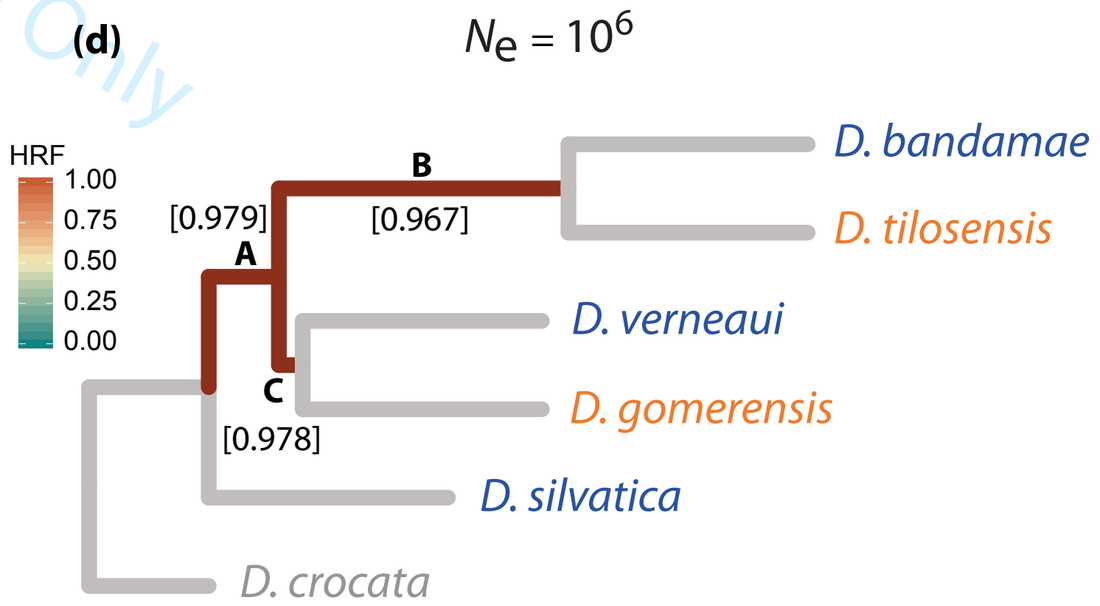

d
D. silvatica
D. verneaui
D. bandamae
Generalists
Specialists
D. gomerensis
D. tilosensis

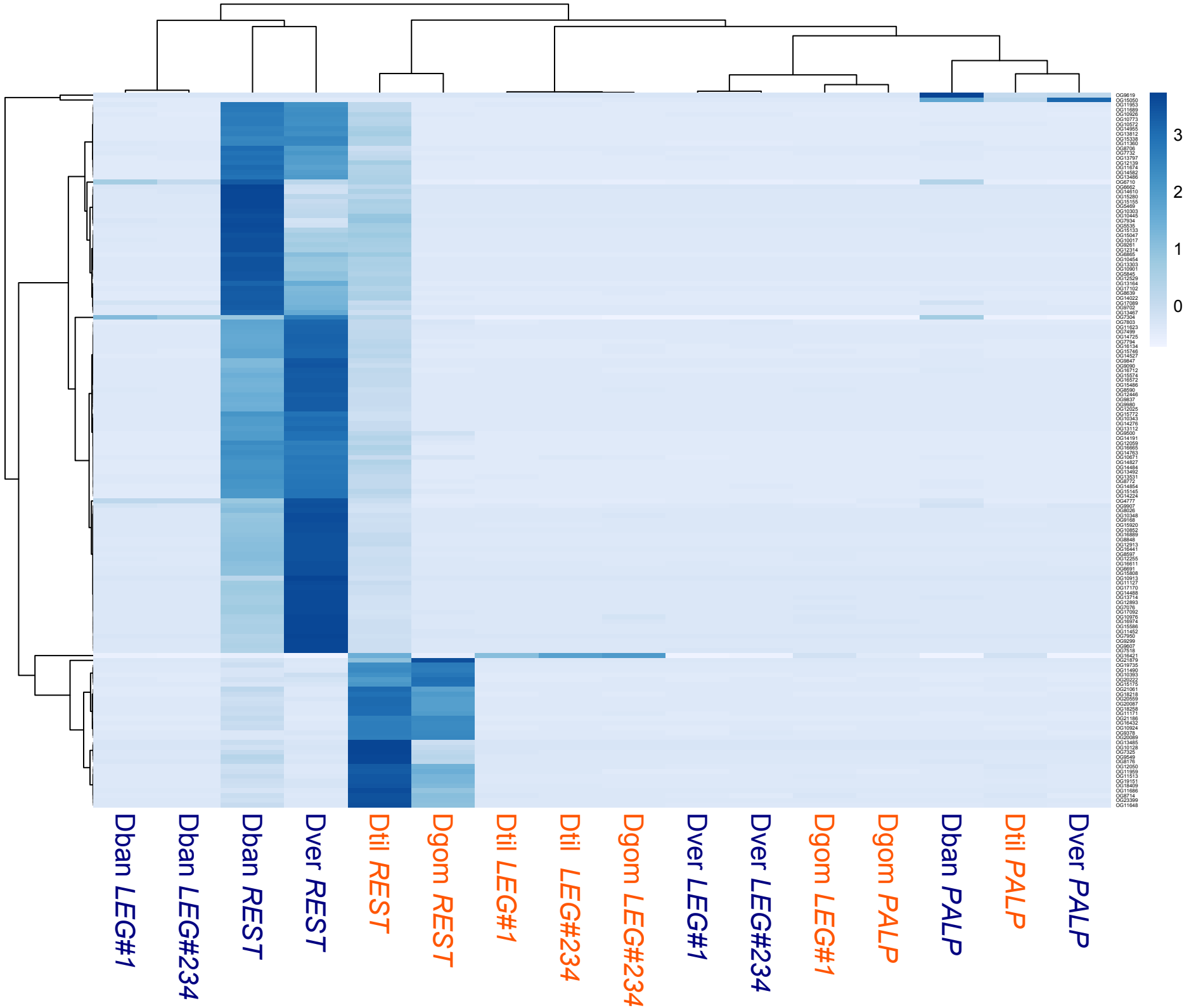


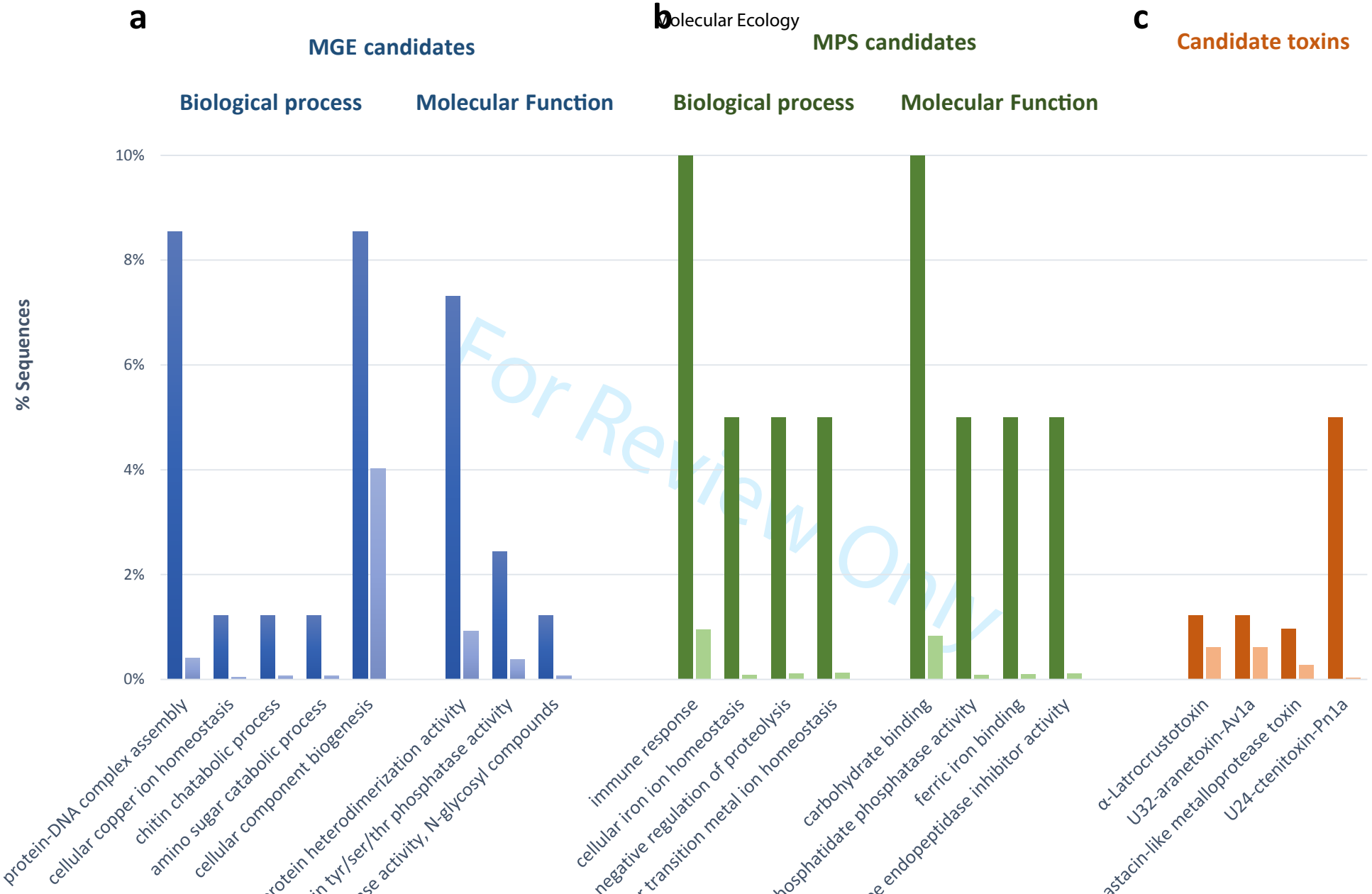



3
2
1
0
Dban LEG#1
Dban LEG#234
Dban REST
Dver REST
Dtil REST
Dgom REST
Dtil LEG#1
Dtil LEG#234
Dgom LEG#234
Dver LEG#1
Dver LEG#234
Dgom LEG#1
Dgom PALP
Dban PALP
Dtil PALP
Dver PALP



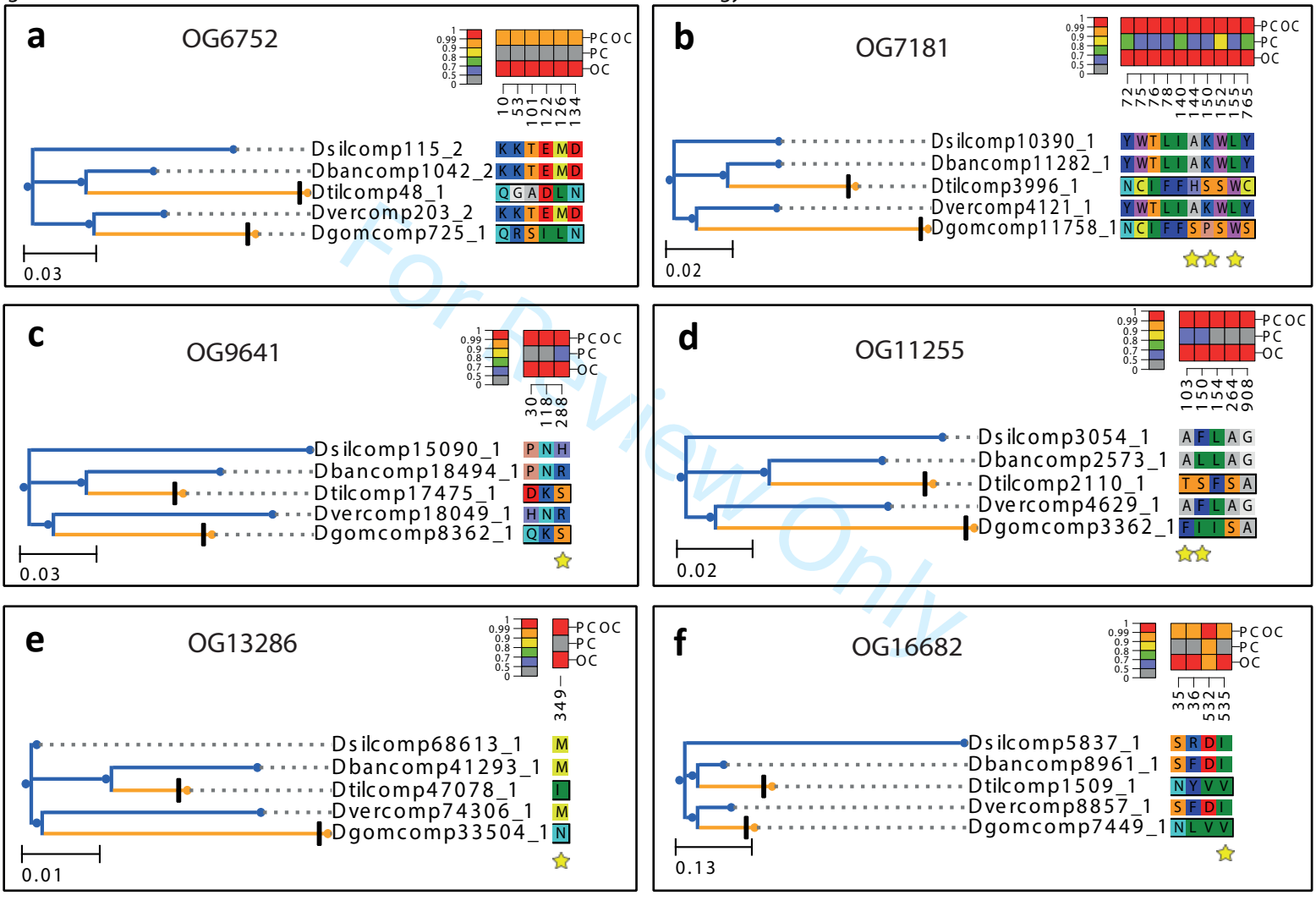


**Table 1. Summary of dietary habits, sampling localities, RNA-seq data and assembly statistics for each surveyed *Dysdera* species**

| | *D. silvatica* | *D. verneaui* | *D. gomerensis* | *D. bandamae* | *D. tilosensis* |
|---|---|---|---|---|---|
| Diet | Generalist | Generalist | Specialist | Generalist | Specialist |
| Locality (in Canary Island) | La Gomera | Tenerife | El Hierro | Gran Canaria | Gran Canaria |
| Total raw reads | 441,835,864 | 527,299,202 | 430,522,240 | 765,653,462 | 678,150,384 |
| Total qualified reads | 418,205,054 | 495,937,054 | 400,095,710 | 746,925,920 | 664,654,842 |
| Transcripts | 236,283 | 441,604 | 213,984 | 296,544 | 316,498 |
| Genes (clustered isoforms) | 170,846 | 347,878 | 177,363 | 221,801 | 229,762 |
| Gene average length (in bp) | 702 | 525 | 622 | 658 | 649 |
| Gene maximum length (in bp) | 26,709 | 27,235 | 27,386 | 27,369 | 25,342 |
| HK genes | 1,136 | 1,194 | 1,232 | 1,153 | 1,159 |
| CEG genes | 807 (457) | 1,180 (457) | 1,111 (457) | 1,033 (457) | 1,143 (457) |
| GO annotated genes | 29,879 | 38,361 | 28,158 | 35,116 | 37,246 |
| Genes with InterPro domain | 30,886 | 40,771 | 29,930 | 37,413 | 39,480 |
| Functional annotated genes[a] | 31,091 | 41,019 | 30,106 | 37,620 | 39,704 |
| Annotated genes[b] | 41,046 | 51,864 | 37,087 | 47,059 | 50,150 |
| Predicted coding sequences (CDS) | 58,966 | 84,114 | 55,914 | 72,352 | 77,756 |
| % not coding genes | 34.51% | 24.18% | 31.53% | 32.62% | 33.84% |
| % not annotated CDS | 69.61% | 61.66% | 66.33% | 65.04% | 64.50% |
| 1to1 orthologs in all species | 9,473 | 9,473 | 9,473 | 9,473 | 9,473 |
| 1to1 orthologs per species pair | - | 19,497 | 19,497 | 24,212 | 24,212 |

[a] GO or Interpro hits.

[b] GO, Interpro or blast hits.